

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2003, 16:12:31 ; Search time 36 Seconds  
(Without alignments)  
25.910 Million cell updates/sec

Title: US-09-733-773B-3

Perfect score: 38

Sequence: 1 YSPSPS 7

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_101002.\*

1:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20:	/SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21:	/SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22:	/SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23:	/SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38	100.0	7	16	AA81828
2	38	100.0	7	17	AA99602
3	38	100.0	7	18	AA18356
4	38	100.0	7	22	AA97327
5	38	100.0	10	22	AA84385
6	38	100.0	10	22	AA84387
7	38	100.0	10	22	AA84391
8	38	100.0	10	22	AA84393
9	38	100.0	171	22	ABG04124
10	38	100.0	288	22	ABG15450

11	38	100.0	1887	22	AB58245
12	35	92.1	7	17	AA99616
13	35	92.1	7	17	AA99608
14	35	92.1	7	17	AA99605
15	35	92.1	7	17	AA99603
16	35	92.1	7	18	AA18371
17	35	92.1	7	18	AA18357
18	35	92.1	7	18	AA18359
19	35	92.1	7	18	AA18362
20	35	92.1	104	21	ABG34182
21	35	92.1	111	22	ABG04114
22	35	92.1	442	22	ABG16776
23	34	89.5	7	16	AA81833
24	34	89.5	7	16	AA81829
25	34	89.5	7	17	AA99615
26	34	89.5	7	17	AA99609
27	34	89.5	7	17	AA99611
28	34	89.5	7	17	AA99612
29	34	89.5	7	18	AA18363
30	34	89.5	7	18	AA18366
31	34	89.5	7	18	AA18367
32	34	89.5	7	18	AA18370
33	34	89.5	84	21	AA852113
34	34	89.5	420	22	AAU30944
35	34	89.5	419	22	AAU30928
36	34	89.5	3005	22	AB58102
37	33	86.8	83	22	AA004309
38	33	86.8	733	19	AAW58572
39	32	84.2	7	17	AA99604
40	32	84.2	7	17	AA99604
41	32	84.2	397	20	AAW18358
42	32	84.2	401	21	AAW90249
43	32	84.2	572	22	ABG25741
44	32	84.2	586	22	ABG25219
45	32	84.2	609	22	ABG17670

## ALIGNMENTS

RESULT 1					
AA81828					
ID	AA81828 standard; Peptide: 7 AA.				
XX	AA81828;				
AC	09-MAY-1996 (first entry)				
XX					
DT	C-terminal domain kinase consensus heptamer.				
XX					
DE	C-terminal domain: nuclear RNA polymerase II; large subunit; kinase.				
XX					
KW	analogue; inhibitor; sporozoan; anti-malarial drug; hyperphosphorylation.				
XX					
OS	Mammalian sp.				
XX					
FH	Key				
FT	Region				
XX	1..7				
XX	/note="repeated upto 52 times"				
XX					
PN	WO9526136-A1.				
XX					
PD	05-OCT-1995.				
XX					
PF	22-MAR-1995; 95WO-US03625.				
XX					
PR	25-MAR-1994; 94US-0218027.				
XX					
PA	(UYDU-) UNTV DUKE.				
XX					
PI	Greenleaf AL, Hardin SE, Lee JM;				
XX					
DR	WPI, 1995-351164/45.				
XX					

Drosophila melanog  
RNA polymerase II  
RNA polymerase II  
RNA polymerase II  
RNA polymerase II  
RNA polymerase II  
RNA polymerase II  
RNA polymerase II  
RNA polymerase II  
RNA polymerase II  
Zea mays protein f  
Novel human diagno  
Novel human diagno  
C-terminal domain  
RNA polymerase II  
RNA polymerase II  
RNA polymerase II  
RNA polymerase II  
RNA polymerase II  
RNA polymerase II  
Human secreted pro  
Novel human diagno  
Novel human diagno  
Drosophila melanog  
Human polyptide  
Tritichoderm reesel  
RNA polymerase II  
Mouse FAST-1 prote  
A forward activin  
Novel human diagno  
Novel human diagno  
Novel human diagno

PT Diagnosis and treatment of parasitic, esp. Plasmodium infections -  
 PT using sporozoan CTD kinase and inhibitors which are distinct from  
 PT mammalian CTD kinase.

XX Disclosure: Page 1; 58pp; English.

CC This peptide is the consensus heptamer of a region designated the  
 CC C-terminal domain (CTD) of the largest subunit of nuclear RNA polymerase  
 CC II. The CTD of humans and other mammals such as mice consists of 52  
 CC repeats of the consensus heptamer, while the CTD of most lower eukaryotes  
 CC consist of fewer repeats of the same consensus. The CTD is essential for  
 CC cell viability and is subject to hyperphosphorylation, which may play an  
 CC important role in initiating transcription and other aspects of RNA  
 CC polymerase II function. The peptide is used to design analogues useful as  
 CC sporozoan CTD kinase inhibitors but incapable of inhibiting mammalian  
 CC CTD kinase.

XX Sequence 7 AA;

Query Match 100.0%; Score 38; DB 16; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 7.8e+05;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 YSPTSPS 7  
 |||||  
 DB 1 YSPTSPS 7

#### RESULT 2

AAR9602  
 ID AAR9602 standard; peptide; 7 AA.

XX AAR9602;

DT 29-JAN-1997 (first entry)

XX RNA polymerase II large subunit C-terminal heptad repeat.

XX RNA polymerase II; heptapeptide repeat; nuclear localisation;  
 XX targeting; delivery; phosphorylation signal.

OS Mammalia.

XX WO9617074-A2.

XX 06-JUN-1996.

XX 01-DEC-1995; 95WO-US15683.

XX 02-DEC-1994; 94US-0348718.

XX (UYVA ) UNIV YALE.

XX Warren SL;

XX WPI; 1996-277787/28.

PT Carboxy-terminal RNA polymerase II peptide(s) - used to deliver  
 PT bioactive agents into discrete compartments in the nucleus of cells  
 XX Claim 3; Page 34; 52pp; English.

XX AAR9602-R99603 are heptad repeat sequences found in the C-terminal of  
 CC the large subunit of mammalian RNA polymerase II. The peptides all  
 CC correspond to a consensus heptad repeat sequence (see AAR9601) found  
 CC not only in mammals but in other eukaryote species e.g. Drosophila,  
 CC Arabidopsis, C. elegans, S. cerevisiae and Plasmodium spp. The heptad  
 CC repeats are used for the delivery of compounds to the nucleus of a cell.  
 CC In vivo the C-terminal heptad repeats are phosphorylated and accumulate  
 CC in discrete subnuclear compartments where pre-mRNA molecules are  
 CC synthesized and spliced. The peptides may be attached to antisense  
 CC oligonucleotides, catalytic RNAs, transgenes, drugs or imaging agents.  
 CC Peptides used for the delivery of agents to the RNA splicing domains

CC within the nucleus are pref. made up of multiple consensus or variable  
 CC repeats.

XX Sequence 7 AA;

Query Match 100.0%; Score 38; DB 17; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 7.8e+05;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 YSPTSPS 7  
 |||||  
 DB 1 YSPTSPS 7

#### RESULT 3

AAW18356  
 ID AAW18356 standard; peptide; 7 AA.

XX AAW18356;

DT 14-JAN-1998 (first entry)

XX RNA polymerase II carboxy-terminus derived peptide, consensus motif.

XX RNA polymerase II carboxy-terminus; pol II CTD; peptide conjugate;  
 XX bioactive molecule; phosphorylation; serine/arginine motif.

XX Homo sapiens.

XX WO9720031-A2.

XX 05-JUN-1997.

XX 29-NOV-1996; 96WO-US19038.

XX 01-DEC-1995; 95US-0566190.

XX (UYVA ) UNIV YALE.

XX Warren SL;

XX WPI; 1997-310586/28.

PT Peptide conjugate for delivery of bioactive compounds to cell  
 PT nucleus - comprises targeting molecule derived from RNA polymerase  
 PT II carboxy-terminus and the bioactive molecule

XX Claim 7; Page 46; 51pp; English.

XX This consensus motif is a specific example of the generic sequence  
 CC AAW18355, which is derived from RNA polymerase II carboxy-terminus  
 CC (pol II CTD). At least two heptapeptide repeats form a peptide conjugate  
 CC with a linker and a bioactive molecule. The peptide conjugate can then  
 CC be used to deliver the bioactive compound, such as a protein, peptide,  
 CC sugar or nucleic acid sequence (e.g. a ribozyme, external guide sequence  
 CC for RNase P, antisense sequence, aptamer, triplex forming  
 CC oligonucleotide, nucleoside, nucleotide, gene, cDNA, mRNA or RNA) to the  
 CC nucleus of a cell. Phosphorylating the peptide alters its association  
 CC with certain molecules in the nucleus, such as proteins having a  
 CC serine/arginine motif and small sub-nuclear ribonucleoprotein (Sm snRNP)  
 CC e.g. phosphorylated pol II CTD (COOH terminal domain) derived peptides  
 CC bind to nuclear proteins associated with transcription and splicing.  
 CC Also, for delivery of molecules which are desired or not desired to be  
 CC in close association with RNA, it may be desirable to phosphorylate or  
 CC leave the peptide unphosphorylated, respectively.

XX Sequence 7 AA;

Query Match 100.0%; Score 38; DB 18; Length 7;  
 Best Local Similarity 100.0%; Pred. No. 7.8e+05;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 YSPTSPS 7

Db 1 YSPTSPS 7

RESULT 4  
AAB97327 standard; peptide; 7 AA.

XX AAB97327;

DT 13-AUG-2001 (first entry)

DE RNA polymerase II epitope peptide.

B cell; toxin; antigen specific; antibody mediated disease; virucide; immunosuppressive; antiinflammatory; antiallergic; antidiabetic; thyromimetic; antithyroid; vasotropic; cardiant; antiulcer; neuroprotective; antirheumatic; antiarthritic; dermatological; ophthalmological; nephrotropic; allergy; autoimmune disorder; skin diseases; autoimmune endocrinopathy; vasculitic syndrome; cardiovascular disease; immunohaematologic disorder; neurologic disease; gastrointestinal disease; collagen vascular disease; renal diseases; pulmonary disease; infertility disorder; RNA polymerase III; systemic sclerosis.

XX Unidentified.

PN WO200132853-A1.

PD 10-MAY-2001.

PF 12-OCT-2000; 2000WO-US28157.

PR 29-OCT-1999; 99US-0162464.

PA (BIOM-) INST APPLIED BIOMEDICINE.

PI Chaplin JW;

DR WPI; 2001-316435/33.

B cell clonal toxin useful for treating autoimmune disorders such as Grave's disease, myocardial infarction, Crohn's disease, multiple sclerosis, comprises a group that causes toxin to be internalized by B cell.

PS Disclosure; Page 32; 46pp; English.

This invention relates to a B cell clonal toxin. The toxin is made from two moieties, the first causes the toxin to be internalised by a B cell, and the second is a biologically acceptable toxin. The invention includes a method for inactivating/killing an antigen specific B cell. A target B cell is contacted with an effective amount of a B cell clonal toxin. The method is useful for selective immunosuppression in conditions characterised by the presence of an unwanted or deleterious immune response, e.g. in the treatment of antigen specific antibody mediated disease conditions. Use of the B cell clonal toxin can result in immunosuppressive, antiinflammatory, antiallergic, virucide, antidiabetic, thyromimetic, antithyroid, vasotropic, cardiant, antiulcer, neuroprotective, antirheumatic, antiarthritic, dermatological, ophthalmological, and nephrotropic activity. The toxin is particularly useful for treating a host suffering from an antigen specific antibody mediated disease condition, where the antigen specific antibody is produced by an antigen-reactive B cell population present in a host. The toxin is useful for treating allergies, viral disease conditions, and autoimmune disorders. Also treated are skin diseases, autoimmune endocrinopathies, vasculitic syndromes, cardiovascular disease, immunohaematologic disorders, gastrointestinal diseases, neurologic diseases, collagen vascular diseases, renal diseases, pulmonary diseases, and infertility disorders. The present sequence represents an RNA polymerase III epitope. An antibody response to this antigen is implicated in systemic sclerosis, a disorder which may be treated using the toxin of the invention.

XX Sequence 7 AA;

Query Match 100.0%; Score 38; DB 22; Length 7;  
Best Local Similarity 100.0%; Pred. No. 7.8e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YSPTSPS 7  
DT 1 YSPTSPS 7

DE RESULT 5

AAAG84385 standard; Peptide; 10 AA.

XX AAAG84385;

DT 11-SEP-2001 (first entry)

DE Arabidopsis thaliana peptide ligand #1025.

XX Plant; peptide pesticide; peptide herbicide; agricultural research.

XX Arabidopsis thaliana.

PN WO200142279-A2.

PD 14-JUN-2001.

PF 13-DEC-2000; 2000WO-GB04781.

PR 13-DEC-1999; 99GB-0029469.

PA (PROT-) PROTEOM LTD.

PI Roberts GW, Heal JR;

DR WPI; 2001-381629/40.

A set of peptide ligands for agricultural research and development of therapeutic agents comprise specific complementary peptides to proteins encoded by genes of plant genomes.

PS Example 4; Page 174; 201pp; English.

The present invention relates to a set of peptide ligands consisting of specific complementary peptides to proteins encoded by genes of plant genomes. The present sequence is one such peptide from Arabidopsis thaliana. The peptides of the present invention are useful in an assay to identify a peptide, especially a peptide pesticide or herbicide. The peptides are also useful for tools for agricultural research and development.

SQ Sequence 10 AA;

Query Match 100.0%; Score 38; DB 22; Length 10;  
Best Local Similarity 100.0%; Pred. No. 2;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YSPTSPS 7  
DT 3 YSPTSPS 9

RESULT 6  
AAAG84387 standard; Peptide; 10 AA.

XX AAAG84387;

DT 11-SEP-2001 (first entry)

DE Arabidopsis thaliana peptide ligand #1027.  
 XX Plant; peptide pesticide; peptide herbicide; agricultural research.  
 XX  
 OS Arabidopsis thaliana.  
 XX  
 PN WO200142279-A2.  
 XX  
 PD 14-JUN-2001.  
 XX  
 PF 13-DEC-2000; 2000WO-GB04781.  
 XX  
 PR 13-DEC-1999; 99GB-0029469.  
 XX  
 PA (PROT-) PROTEOM LTD.  
 XX  
 PI Roberts GW, Heal JR;  
 XX  
 DR WPI; 2001-381629/40.  
 XX  
 PS Example 4; Page 174; 201pp; English.  
 XX  
 CC The present invention relates to a set of peptide ligands consisting of  
 CC specific complementary peptides to proteins encoded by genes of plant  
 CC genomes. The present sequence is one such peptide from Arabidopsis  
 CC thaliana. The peptides of the present invention are useful in an assay to  
 CC identify a peptide, especially a peptide pesticide or herbicide. The  
 CC peptides are also useful for tools for agricultural research and  
 CC development.  
 CC  
 SQ Sequence 10 AA;  
 XX  
 Query Match 100.0%; Score 38; DB 22; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 2;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 YSPSPS 7  
 |||||  
 Db 2 YSPSPS 8  
 XX  
 RESULT 7  
 AAG84391  
 ID AAG84391 standard; Peptide; 10 AA.  
 XX  
 AC AAG84391;  
 XX  
 DT 11-SEP-2001 (first entry)  
 XX  
 DE Arabidopsis thaliana peptide ligand #1031.  
 XX  
 KW Plant; peptide pesticide; peptide herbicide; agricultural research.  
 XX  
 OS Arabidopsis thaliana.  
 XX  
 PN WO200142279-A2.  
 XX  
 PD 14-JUN-2001.  
 XX  
 PF 13-DEC-2000; 2000WO-GB04781.  
 XX  
 PR 13-DEC-1999; 99GB-0029469.  
 XX  
 PA (PROT-) PROTEOM LTD.  
 XX  
 PI Roberts GW, Heal JR;  
 XX  
 DR WPI; 2001-381629/40.  
 XX

PT A set of peptide ligands for agricultural research and development of  
 PT therapeutic agents comprise specific complementary peptides to proteins  
 PT encoded by genes of plant genomes -  
 XX  
 PS Example 4; Page 175; 201pp; English.  
 XX  
 CC The present invention relates to a set of peptide ligands consisting of  
 CC specific complementary peptides to proteins encoded by genes of plant  
 CC genomes. The present sequence is one such peptide from Arabidopsis  
 CC thaliana. The peptides of the present invention are useful in an assay to  
 CC identify a peptide, especially a peptide pesticide or herbicide. The  
 CC peptides are also useful for tools for agricultural research and  
 CC development.  
 CC  
 SQ Sequence 10 AA;  
 XX  
 Query Match 100.0%; Score 38; DB 22; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 2;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 YSPSPS 7  
 |||||  
 Db 3 YSPSPS 9  
 XX  
 RESULT 8  
 AAG84393  
 ID AAG84393 standard; Peptide; 10 AA.  
 XX  
 AC AAG84393;  
 XX  
 DT 11-SEP-2001 (first entry)  
 XX  
 DE Arabidopsis thaliana peptide ligand #1033.  
 XX  
 KW Plant; peptide pesticide; peptide herbicide; agricultural research.  
 XX  
 OS Arabidopsis thaliana.  
 XX  
 PN WO200142279-A2.  
 XX  
 PD 14-JUN-2001.  
 XX  
 PF 13-DEC-2000; 2000WO-GB04781.  
 XX  
 PR 13-DEC-1999; 99GB-0029469.  
 XX  
 PA (PROT-) PROTEOM LTD.  
 XX  
 PI Roberts GW, Heal JR;  
 XX  
 DR WPI; 2001-381629/40.  
 XX  
 PT A set of peptide ligands for agricultural research and development of  
 PT therapeutic agents comprise specific complementary peptides to proteins  
 PT encoded by genes of plant genomes -  
 XX  
 PS Example 4; Page 175; 201pp; English.  
 XX  
 CC The present invention relates to a set of peptide ligands consisting of  
 CC specific complementary peptides to proteins encoded by genes of plant  
 CC genomes. The present sequence is one such peptide from Arabidopsis  
 CC thaliana. The peptides of the present invention are useful in an assay to  
 CC identify a peptide, especially a peptide pesticide or herbicide. The  
 CC peptides are also useful for tools for agricultural research and  
 CC development.  
 CC  
 SQ Sequence 10 AA;  
 XX  
 Query Match 100.0%; Score 38; DB 22; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 2;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YSPTSPS 7  
|||||||  
Db 2 YSPTSPS 8

RESULT 9  
ABG04124  
ID ABG04124 standard; Protein: 171 AA.  
XX  
XX ABG04124;  
DT 13-FEB-2002 (first entry)  
XX  
XX Novel human diagnostic protein #4115.  
DE  
XX  
XX Human: chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
XX Homo sapiens.  
OS  
XX WO200175067-A2.  
PN  
XX 11-OCT-2001.  
PD  
XX 30-MAR-2001; 2001WO-US08631.  
PF  
XX 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
XX (HYSE-) HYSEQ INC.  
PA  
XX  
XX Drmanac RT, Liu C, Tang YT;  
PI  
XX  
XX WPI: 2001-639362/73.  
DR N-PSDB; AAS68311.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
XX  
PS Claim 20; SEQ ID No 34483; 103bp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX  
SQ Sequence 171 AA;

Query Match 100.0%; Score 38; DB 22; Length 171;  
Best Local Similarity 100.0%; Pred. No. 40;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YSPTSPS 7  
|||||||

Db 119 YSPTSPS 125

RESULT 10  
ABG15450  
ID ABG15450 standard; Protein: 288 AA.  
XX  
XX ABG15450;  
DT 18-FEB-2002 (first entry)  
XX  
XX Novel human diagnostic protein #15441.  
DE  
XX  
XX Human: chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
XX Homo sapiens.  
OS  
XX WO200175067-A2.  
PN  
XX 11-OCT-2001.  
PD  
XX 30-MAR-2001; 2001WO-US08631.  
PF  
XX 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
XX (HYSE-) HYSEQ INC.  
PA  
XX  
XX Drmanac RT, Liu C, Tang YT;  
PI  
XX  
XX WPI: 2001-639362/73.  
DR N-PSDB; AAS79637.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
XX  
PS Claim 20; SEQ ID No 45809; 103bp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX  
SQ Sequence 288 AA;

Query Match 100.0%; Score 38; DB 22; Length 288;  
Best Local Similarity 100.0%; Pred. No. 69;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YSPTSPS 7  
|||||||

Db 71 YSPTSPS 77

RESULT 11  
ID ABB58245 standard; Protein; 1887 AA.  
XX ABB58245;  
AC ABB58245;  
XX 26-MAR-2002 (first entry)  
DT 26-MAR-2002 (first entry)  
XX Drosophila melanogaster polypeptide SEQ ID NO 1527.  
DE Drosophila melanogaster polypeptide SEQ ID NO 1527.  
XX Drosophila melanogaster polypeptide SEQ ID NO 1527.  
KW Drosophila melanogaster polypeptide SEQ ID NO 1527.  
XX Drosophila melanogaster polypeptide SEQ ID NO 1527.  
OS Drosophila melanogaster.  
XX WO200171042-A2.  
PN WO200171042-A2.  
XX 27-SEP-2001.  
PD 27-SEP-2001.  
XX 23-MAR-2001; 2001WO-US09231.  
PF 23-MAR-2001; 2000US-191637P.  
XX 23-MAR-2001; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX (PEKE ) PE CORP NY.  
PA (PEKE ) PE CORP NY.  
XX Venter JC, Adams M, Li PMD, Myers EW;  
PI WPI; 2001-656860/75.  
XX N-PSDB; ABL02348.  
DR N-PSDB; ABL02348.  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
interactions -  
XX Interactions -  
XX Disclosure; SEQ ID NO 1527; 21pp + Sequence Listing; English.  
PS The invention relates to an isolated nucleic acid detection reagent  
XX capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA  
CC sequences (AB101840-AB16175) and the encoded proteins  
CC (AB87737-AB872072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX Sequence 1887 AA;  
SQ Sequence 1887 AA;  
Query Match 100.0%; Score 38; DB 22; Length 1887;  
Best Local Similarity 100.0%; Pred. No. 5e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 YSPSPS 7  
DB 1685 YSPSPS 1691  
RESULT 12  
ID AAR99616 standard; peptide; 7 AA.  
XX AAR99616;  
AC AAR99616;  
XX 29-JAN-1997 (first entry)  
DT 29-JAN-1997 (first entry)  
XX RNA polymerase II large subunit C-terminal heptad repeat.  
DE RNA polymerase II large subunit C-terminal heptad repeat.  
XX RNA polymerase II; heptapeptide repeat; nuclear localisation;  
KW targeting; delivery; phosphorylation signal.  
KW targeting; delivery; phosphorylation signal.

XX Mammalia.  
OS Mammalia.  
XX WO9617074-A2.  
PN WO9617074-A2.  
XX 06-JUN-1996.  
PD 06-JUN-1996.  
XX 01-DEC-1995; 95WO-US15683.  
PF 01-DEC-1995; 95WO-US15683.  
XX 02-DEC-1994; 94US-0348718.  
PR 02-DEC-1994; 94US-0348718.  
XX (UYVA ) UNIV YALE.  
PA (UYVA ) UNIV YALE.  
XX Warren SL;  
PI WPI; 1996-277787/28.  
DR WPI; 1996-277787/28.  
XX Carboxy-terminal RNA polymerase II peptide(s) - used to deliver  
PT bioactive agents into discrete compartments in the nucleus of cells  
XX bioactive agents into discrete compartments in the nucleus of cells  
XX Claim 3; Page 34; 52pp; English.  
PS AAR9602-R99603 are heptad repeat sequences found in the C-terminal of  
XX the large subunit of mammalian RNA polymerase II. The peptides all  
XX correspond to a consensus heptad repeat sequence (see AAR9601) found  
CC not only in mammals but in other eukaryote species e.g. Drosophila,  
CC Arabidopsis, C. elegans, S. cerevisiae and Plasmodium spp. The heptad  
CC repeats are used for the delivery of compounds to the nucleus of a cell.  
CC In vivo the C-terminal heptad repeats are phosphorylated and accumulate  
CC in discrete subnuclear compartments where pre-mRNA molecules are  
CC synthesized and spliced. The peptides may be attached to antisense  
CC oligonucleotides, catalytic RNAs, transgenes, drugs or imaging agents.  
CC Peptides used for the delivery of agents to the RNA splicing domains  
CC within the nucleus are pref. made up of multiple consensus or variable  
XX repeats.  
XX Sequence 7 AA;  
SQ Sequence 7 AA;  
Query Match 92.1%; Score 35; DB 17; Length 7;  
Best Local Similarity 85.7%; Pred. No. 7.8e+05;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
OY 1 YSPSPS 7  
DB 1 YTPSPS 7  
RESULT 13  
ID AAR9608 standard; peptide; 7 AA.  
XX AAR9608;  
AC AAR9608;  
XX 29-JAN-1997 (first entry)  
DT 29-JAN-1997 (first entry)  
XX RNA polymerase II large subunit C-terminal heptad repeat.  
DE RNA polymerase II large subunit C-terminal heptad repeat.  
XX RNA polymerase II; heptapeptide repeat; nuclear localisation;  
KW targeting; delivery; phosphorylation signal.  
XX Mammalia.  
OS Mammalia.  
XX WO9617074-A2.  
PN WO9617074-A2.  
XX 06-JUN-1996.  
PD 06-JUN-1996.  
XX 01-DEC-1995; 95WO-US15683.  
PF 01-DEC-1995; 95WO-US15683.  
XX 02-DEC-1994; 94US-0348718.  
PR 02-DEC-1994; 94US-0348718.  
XX (UYVA ) UNIV YALE.  
PA (UYVA ) UNIV YALE.  
XX Warren SL;  
PI Warren SL;

XX DR WPI; 1996-277787/28.  
 XX PT Carboxy-terminal RNA polymerase II peptide(s) - used to deliver  
 PT bioactive agents into discrete compartments in the nucleus of cells  
 XX PS Claim 3; Page 34; 52pp; English.  
 XX CC AAR99602-R99603 are heptad repeat sequences found in the C-terminal of  
 CC the large subunit of mammalian RNA polymerase II. The peptides all  
 CC correspond to a consensus heptad repeat sequence (see AAR99601) found  
 CC not only in mammals but in other eukaryote species e.g. Drosophila,  
 CC Arabidopsis, C. elegans, S. cerevisiae and Plasmodium spp. The heptad  
 CC repeats are used for the delivery of compounds to the nucleus of a cell.  
 CC In vivo the C-terminal heptad repeats are phosphorylated and accumulate  
 CC in discrete subnuclear compartments where pre-mRNA molecules are  
 CC synthesized and spliced. The peptides may be attached to antisense  
 CC oligonucleotides, catalytic RNAs, transgenes, drugs or imaging agents.  
 CC Peptides used for the delivery of agents to the RNA splicing domains  
 CC within the nucleus are pref. made up of multiple consensus or variable  
 CC repeats.  
 CC SQ Sequence 7 AA;  
 CC  
 CC Query Match 92.1%; Score 35; DB 17; Length 7;  
 CC Best Local Similarity 85.7%; Pred. No. 7.8e+05;  
 CC Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 CC  
 CC QY 1 YSP7SPS 7  
 CC |||||:  
 CC 1 YSP7SP7 7  
 CC Db  
 CC  
 CC RESULT 14  
 CC AAR99605  
 CC ID AAR99605 standard; peptide; 7 AA.  
 CC XX  
 CC AC AAR99605;  
 CC XX  
 CC DT 29-JAN-1997 (first entry)  
 CC XX  
 CC DE RNA polymerase II large subunit C-terminal heptad repeat.  
 CC XX  
 CC KM RNA polymerase II; heptapeptide repeat; nuclear localisation;  
 CC KW targeting; delivery; phosphorylation signal.  
 CC XX  
 CC OS Mammalia.  
 CC XX  
 CC PN WO9617074-A2.  
 CC PD 06-JUN-1996.  
 CC XX  
 CC PF 01-DEC-1995; 95WO-US15683.  
 CC XX  
 CC PR 02-DEC-1994; 94US-0348718.  
 CC XX  
 CC PA (UYA ) UNIV YALE.  
 CC XX  
 CC PI Warren SL;  
 CC XX  
 CC DR WPI; 1996-277787/28.  
 CC XX  
 CC PT Carboxy-terminal RNA polymerase II peptide(s) - used to deliver  
 PT bioactive agents into discrete compartments in the nucleus of cells  
 XX PS Claim 3; Page 34; 52pp; English.  
 XX CC AAR99602-R99603 are heptad repeat sequences found in the C-terminal of  
 CC the large subunit of mammalian RNA polymerase II. The peptides all  
 CC correspond to a consensus heptad repeat sequence (see AAR99601) found  
 CC not only in mammals but in other eukaryote species e.g. Drosophila,  
 CC Arabidopsis, C. elegans, S. cerevisiae and Plasmodium spp. The heptad  
 CC repeats are used for the delivery of compounds to the nucleus of a cell.

CC CC In vivo the C-terminal heptad repeats are phosphorylated and accumulate  
 CC in discrete subnuclear compartments where pre-mRNA molecules are  
 CC synthesized and spliced. The peptides may be attached to antisense  
 CC oligonucleotides, catalytic RNAs, transgenes, drugs or imaging agents.  
 CC Peptides used for the delivery of agents to the RNA splicing domains  
 CC within the nucleus are pref. made up of multiple consensus or variable  
 CC repeats.  
 CC SQ Sequence 7 AA;  
 CC  
 CC Query Match 92.1%; Score 35; DB 17; Length 7;  
 CC Best Local Similarity 85.7%; Pred. No. 7.8e+05;  
 CC Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 CC  
 CC QY 1 YSP7SPS 7  
 CC |||||:  
 CC 1 YSP7SPA 7  
 CC Db  
 CC  
 CC RESULT 15  
 CC AAR99603  
 CC ID AAR99603 standard; peptide; 7 AA.  
 CC XX  
 CC AC AAR99603;  
 CC XX  
 CC DT 29-JAN-1997 (first entry)  
 CC XX  
 CC DE RNA polymerase II large subunit C-terminal heptad repeat.  
 CC XX  
 CC KM RNA polymerase II; heptapeptide repeat; nuclear localisation;  
 CC KW targeting; delivery; phosphorylation signal.  
 CC XX  
 CC OS Mammalia.  
 CC XX  
 CC PN WO9617074-A2.  
 CC PD 06-JUN-1996.  
 CC XX  
 CC PF 01-DEC-1995; 95WO-US15683.  
 CC XX  
 CC PR 02-DEC-1994; 94US-0348718.  
 CC XX  
 CC PA (UYA ) UNIV YALE.  
 CC XX  
 CC PI Warren SL;  
 CC XX  
 CC DR WPI; 1996-277787/28.  
 CC XX  
 CC PT Carboxy-terminal RNA polymerase II peptide(s) - used to deliver  
 PT bioactive agents into discrete compartments in the nucleus of cells  
 XX PS Claim 3; Page 34; 52pp; English.  
 XX CC AAR99602-R99603 are heptad repeat sequences found in the C-terminal of  
 CC the large subunit of mammalian RNA polymerase II. The peptides all  
 CC correspond to a consensus heptad repeat sequence (see AAR99601) found  
 CC not only in mammals but in other eukaryote species e.g. Drosophila,  
 CC Arabidopsis, C. elegans, S. cerevisiae and Plasmodium spp. The heptad  
 CC repeats are used for the delivery of compounds to the nucleus of a cell.  
 CC In vivo the C-terminal heptad repeats are phosphorylated and accumulate  
 CC in discrete subnuclear compartments where pre-mRNA molecules are  
 CC synthesized and spliced. The peptides may be attached to antisense  
 CC oligonucleotides, catalytic RNAs, transgenes, drugs or imaging agents.  
 CC Peptides used for the delivery of agents to the RNA splicing domains  
 CC within the nucleus are pref. made up of multiple consensus or variable  
 CC repeats.  
 CC SQ Sequence 7 AA;  
 CC  
 CC Query Match 92.1%; Score 35; DB 17; Length 7;  
 CC Best Local Similarity 85.7%; Pred. No. 7.8e+05;  
 CC Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Mon Jun 23 16:45:01 2003

us-09-733-773b-3.rag

Page 8

Oy 1 YSPRSPS 7  
|||||:  
Db 1 YSPRSPN 7

Search completed: June 23, 2003, 16:17:20  
Job time : 37 secs

---



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2003, 16:16:46 ; Search time 14 Seconds  
(without alignments)  
14.711 Million cell updates/sec

Title: US-09-733-773B-3

Perfect score: 38

Sequence: 1 YSPSPS 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents-AA:\*

1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep:\*

2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep:\*

3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep:\*

4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep:\*

5: /cgn2\_6/ptodata/1/1aa/PTCUS.COMB.pep:\*

6: /cgn2\_6/ptodata/1/1aa/Backfill1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	38	100.0	7	1 US-08-218-027A-1	Sequence 1, Appl
2	38	100.0	7	1 US-08-040-548-40	Sequence 40, Appl
3	38	100.0	7	1 US-08-466-344-40	Sequence 40, Appl
4	38	100.0	7	3 US-08-348-718-1	Sequence 1, Appl
5	38	100.0	7	3 US-08-566-190-1	Sequence 1, Appl
6	38	100.0	31	3 US-09-280-047-10	Sequence 10, Appl
7	38	100.0	31	4 US-08-208-573B-10	Sequence 10, Appl
8	35	92.1	7	3 US-08-348-718-2	Sequence 2, Appl
9	35	92.1	7	3 US-08-566-190-2	Sequence 2, Appl
10	35	92.1	7	3 US-08-566-190-4	Sequence 4, Appl
11	35	92.1	7	3 US-08-566-190-7	Sequence 7, Appl
12	35	92.1	7	3 US-08-566-190-16	Sequence 16, Appl
13	35	92.1	7	3 US-08-218-027A-2	Sequence 2, Appl
14	34	89.5	7	1 US-08-218-027A-6	Sequence 6, Appl
15	34	89.5	7	3 US-08-348-718-3	Sequence 3, Appl
16	34	89.5	7	3 US-08-566-190-8	Sequence 8, Appl
17	34	89.5	7	3 US-08-566-190-11	Sequence 11, Appl
18	34	89.5	7	3 US-08-566-190-12	Sequence 12, Appl
19	34	89.5	7	3 US-08-566-190-15	Sequence 15, Appl
20	32	84.2	7	3 US-08-566-190-3	Sequence 3, Appl
21	32	84.2	397	4 US-09-087-134-17	Sequence 17, Appl
22	31	81.6	7	1 US-08-218-027A-3	Sequence 3, Appl
23	31	81.6	7	1 US-08-218-027A-4	Sequence 4, Appl
24	31	81.6	7	1 US-08-218-027A-7	Sequence 7, Appl
25	31	81.6	7	3 US-08-566-190-9	Sequence 9, Appl
26	31	81.6	7	3 US-08-566-190-10	Sequence 10, Appl
27	31	81.6	7	3 US-08-566-190-14	Sequence 14, Appl

28	31	81.6	46	3 US-08-856-074A-39	Sequence 39, Appl
29	31	81.6	60	3 US-08-856-074A-1	Sequence 1, Appl
30	31	81.6	147	4 US-09-134-001C-3180	Sequence 3180, Ap
31	31	81.6	180	1 US-08-447-010-6	Sequence 6, Appl
32	31	81.6	282	5 PCT-US94-09752-3	Sequence 3, Appl
33	31	81.6	327	2 US-08-907-674-3	Sequence 3, Appl
34	31	81.6	327	2 US-09-215-087-3	Sequence 3, Appl
35	31	81.6	327	2 US-09-391-959-3	Sequence 3, Appl
36	31	81.6	654	4 US-09-315-127-11	Sequence 11, Appl
37	31	81.6	654	4 US-09-315-127-12	Sequence 12, Appl
38	31	81.6	739	4 US-09-022-983-2	Sequence 2, Appl
39	30	78.9	7	3 US-08-566-190-17	Sequence 17, Appl
40	30	78.9	7	3 US-08-566-190-19	Sequence 19, Appl
41	29	76.3	7	3 US-08-566-190-5	Sequence 5, Appl
42	29	76.3	92	4 US-09-334-601-30	Sequence 30, Appl
43	29	76.3	205	2 US-08-912-227-4	Sequence 4, Appl
44	29	76.3	205	4 US-09-513-584-4	Sequence 4, Appl
45	29	76.3	224	1 US-08-173-510B-87	Sequence 87, Appl

## ALIGNMENTS

RESULT 1  
US-08-218-027A-1  
Sequence 1, Application US/08218027A  
Patent No. 5565327  
GENERAL INFORMATION:  
APPLICANT: Greenleaf, Arno L.  
APPLICANT: Lee, Jae M.  
TITLE OF INVENTION: METHODS OF DIAGNOSING PARASITIC  
TITLE OF INVENTION: INFECTIONS AND OF TESTING DRUG SUSCEPTIBILITY OF PARASITES  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Kenneth D. Sibley  
STREET: P.O. Drawer 34009  
CITY: Charlotte  
STATE: NC. 5565327th Carolina  
COUNTRY: USA  
ZIP: 28234  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/218, 027A  
FILING DATE: 25-MAR-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Sibley, Kenneth D.  
REGISTRATION NUMBER: 31,665  
REFERENCE/DOCKET NUMBER: 5405-91  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-420-2200  
TELEFAX: 919-881-3175  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-218-027A-1  
Query Match 100.0%; Score 38; DB 1; Length 7;  
Best Local Similarity 100.0%; Pred. No. 2e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 YSPSPS 7  
|||||||  
Db 1 YSPSPS 7

RESULT 2  
US-08-040-548-40  
; Sequence 40, Application US/08040548  
; Patent No. 5763209  
; GENERAL INFORMATION:  
; APPLICANT: Subhane, Vikas P.  
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO THE  
; TITLE OF INVENTION: FUNCTIONAL DOMAINS OF DNA BINDING PROTEINS  
; NUMBER OF SEQUENCES: 67  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Arnold, White & Durkee  
; STREET: 321 No. 5763209th Clark Street, Suite 800  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: U.S.A.  
; ZIP: 60610  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/040,548  
; FILING DATE:  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coughlin, Daniel F.  
; REGISTRATION NUMBER: 36,111  
; REFERENCE/DOCKET NUMBER: arc0067  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (312) 744-0090  
; TELEFAX: (312) 245-4961  
; INFORMATION FOR SEQ ID NO: 40:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; US-08-040-548-40

Query Match 100.0%; Score 38; DB 1; Length 7;  
Best Local Similarity 100.0%; Pred. No. 2e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Caps 0;

QY 1 YSPTSPS 7  
| | | | |  
Db 1 YSPTSPS 7

RESULT 3  
US-08-466-344-40  
; Sequence 40, Application US/08466344  
; Patent No. 5773583  
; GENERAL INFORMATION:  
; APPLICANT: Subhane, Vikas P.  
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO THE  
; TITLE OF INVENTION: FUNCTIONAL DOMAINS OF DNA BINDING PROTEINS  
; NUMBER OF SEQUENCES: 67  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Arnold, White & Durkee  
; STREET: 321 No. 5773583th Clark Street, Suite 800  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: U.S.A.  
; ZIP: 60610  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/466,344  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/040,548  
; FILING DATE: 31-MAR-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coughlin, Daniel F.  
; REGISTRATION NUMBER: 36,111  
; REFERENCE/DOCKET NUMBER: arc0067  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (312) 744-0090  
; TELEFAX: (312) 245-4961  
; INFORMATION FOR SEQ ID NO: 40:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; US-08-466-344-40

Query Match 100.0%; Score 38; DB 1; Length 7;  
Best Local Similarity 100.0%; Pred. No. 2e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Caps 0;

QY 1 YSPTSPS 7  
| | | | |  
Db 1 YSPTSPS 7

RESULT 4  
US-08-348-718-1  
; Sequence 1, Application US/08348718  
; Patent No. 6007985  
; GENERAL INFORMATION:  
; APPLICANT: Stephen L. Warren  
; TITLE OF INVENTION: RNA Polymerase II Antibodies and  
; TITLE OF INVENTION: Macromolecule Delivery Method  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Patrea L. Pabst  
; STREET: 2800 One Atlantic Center  
; STREET: 1201 West Peachtree Street  
; CITY: Atlanta  
; STATE: Georgia  
; COUNTRY: USA  
; ZIP: 30309-3450  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/348,718  
; FILING DATE: December 2, 1994  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/098,991  
; FILING DATE: July 8, 1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/955,338  
; FILING DATE: October 1, 1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Pabst, Patrea L.  
; REGISTRATION NUMBER: 31,284  
; REFERENCE/DOCKET NUMBER: YU106CIP(2)  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (404) 873-8795  
; TELEFAX: (404) 873-8795  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:

LENGTH: 7 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-348-718-1

Query Match 100.0%; Score 38; DB 3; Length 7;  
Best Local Similarity 100.0%; Pred. No. 2e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 YSPSPS 7  
1111111  
DB 1 YSPSPS 7

RESULT 5  
US-08-566-190-1  
Sequence 1, Application US/08566190  
Patent No. 6090784

GENERAL INFORMATION:  
APPLICANT: Warren, Stephen L.  
TITLE OF INVENTION: RNA Polymerase II Peptides and Methods of Use  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESS: Patrea L. Pabst  
STREET: 2800 One Atlantic Center, 1201 West Peachtree Street  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: USA  
ZIP: 30309-3450

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/566,190  
FILING DATE: 1-DEC-1995  
CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:  
NAME: Pabst, Patrea L.  
REGISTRATION NUMBER: 31,284  
REFERENCE/DOCKET NUMBER: YU116  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (404) 873-8794  
TELEFAX: (404) 873-8795

INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear

MOLECULE TYPE: peptide  
US-08-566-190-1

Query Match 100.0%; Score 38; DB 3; Length 7;  
Best Local Similarity 100.0%; Pred. No. 2e+05;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 YSPSPS 7  
1111111  
DB 1 YSPSPS 7

RESULT 6  
US-09-280-047-10  
Sequence 10, Application US/09280047  
Patent No. 6066462

GENERAL INFORMATION:  
APPLICANT: GODELL, SAID A.  
TITLE OF INVENTION: QUANTITATION OF INDIVIDUAL PROTEIN  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Ross & Stevens, S.C.  
STREET: One South Pinckney Street, P.O. Box 2599  
CITY: Madison  
STATE: Wisconsin  
COUNTRY: U.S.A.  
ZIP: 53701-2599

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/280,047  
FILING DATE:

CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/208,573  
FILING DATE: 04-MAR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: SARA, CHARLES S.

REGISTRATION NUMBER: 30,492  
REFERENCE/DOCKET NUMBER: 34506.006  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (608) 257-5353  
TELEFAX: (608) 257-9175  
INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:  
LENGTH: 31 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-09-280-047-10

Query Match 100.0%; Score 38; DB 3; Length 31;  
Best Local Similarity 100.0%; Pred. No. 1.6;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 YSPSPS 7  
1111111  
DB 4 YSPSPS 10

RESULT 7  
US-08-208-573b-10  
Sequence 10, Application US/08208573B  
Patent No. 6348310

GENERAL INFORMATION:  
APPLICANT: GODELL, SAID A.  
TITLE OF INVENTION: QUANTITATION OF INDIVIDUAL PROTEIN  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Ross & Stevens, S.C.  
STREET: One South Pinckney Street, P.O. Box 2599  
CITY: Madison  
STATE: Wisconsin  
COUNTRY: U.S.A.  
ZIP: 53701-2599

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/208,573B  
FILING DATE: 10-DEC-1998  
CLASSIFICATION: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: SARA, CHARLES S.

REGISTRATION NUMBER: 30,492  
REFERENCE/DOCKET NUMBER: 34506.006  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (608) 257-5353

TELEFAX: (608) 257-9175  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 31 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 10:  
US-08-208-573b-10

Query Match 100.0%; Score 38; DB 4; Length 31;  
Best Local Similarity 100.0%; Pred. No. 1.6;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YSPTSPS 7  
DB 4 YSPTSPS 10

RESULT 8  
US-08-348-718-2  
Sequence 2, Application US/08348718  
Patent No. 6007985  
GENERAL INFORMATION:  
APPLICANT: Stephen L. Warren  
TITLE OF INVENTION: RNA Polymerase II Antibodies and  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Patrea L. Pabst  
STREET: 2800 One Atlantic Center  
STREET: 1201 West Peachtree Street  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: USA  
ZIP: 30309-3450  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/348,718  
FILING DATE: December 2, 1994  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/098,991  
FILING DATE: July 8, 1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/955,338  
FILING DATE: October 1, 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Pabst, Patrea L.  
REGISTRATION NUMBER: 31,284  
REFERENCE/DOCKET NUMBER: YU10CIP(2)  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (404) 873-8794  
TELEFAX: (404) 873-8795  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-348-718-2

Query Match 92.1%; Score 35; DB 3; Length 7;  
Best Local Similarity 85.7%; Pred. No. 2e+05;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 YSPTSPS 7  
DB 1 YSPTSPA 7

RESULT 9  
US-08-566-190-2  
Sequence 2, Application US/08566190  
Patent No. 6090784  
GENERAL INFORMATION:  
APPLICANT: Warren, Stephen L.  
TITLE OF INVENTION: RNA Polymerase II Peptides and Methods of Use  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Patrea L. Pabst  
STREET: 2800 One Atlantic Center, 1201 West Peachtree Street  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: USA  
ZIP: 30309-3450  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/566,190  
FILING DATE: 1-DEC-1995  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Pabst, Patrea L.  
REGISTRATION NUMBER: 31,284  
REFERENCE/DOCKET NUMBER: YU116  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (404) 873-8794  
TELEFAX: (404) 873-8795  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-566-190-2

Query Match 92.1%; Score 35; DB 3; Length 7;  
Best Local Similarity 85.7%; Pred. No. 2e+05;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 YSPTSPS 7  
DB 1 YSPTSPN 7

RESULT 10  
US-08-566-190-4  
Sequence 4, Application US/08566190  
Patent No. 6090784  
GENERAL INFORMATION:  
APPLICANT: Warren, Stephen L.  
TITLE OF INVENTION: RNA Polymerase II Peptides and Methods of Use  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Patrea L. Pabst  
STREET: 2800 One Atlantic Center, 1201 West Peachtree Street  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: USA  
ZIP: 30309-3450  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/566,190  
FILING DATE: 1-DEC-1995

CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Pabst, Patrea L.  
REGISTRATION NUMBER: 31,284  
REFERENCE/DOCKET NUMBER: YU116  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (404) 873-8794  
TELEFAX: (404) 873-8795  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-566-190-4

Query Match 92.1%; Score 35; DB 3; Length 7;  
Best Local Similarity 85.7%; Pred. No. 2e+05;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 YSPTSPS 7  
Db 1 YSPTSPA 7

RESULT 11  
US-08-566-190-7  
Sequence 7, Application US/08566190  
Patent No. 6090784  
GENERAL INFORMATION:  
APPLICANT: Warren, Stephen L.  
TITLE OF INVENTION: RNA Polymerase II Peptides and Methods of Use  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Patrea L. Pabst  
STREET: 2800 One Atlantic Center, 1201 West Peachtree Street  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: USA  
ZIP: 30309-3450  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/566,190  
FILING DATE: 1-DEC-1995  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Pabst, Patrea L.  
REGISTRATION NUMBER: 31,284  
REFERENCE/DOCKET NUMBER: YU116  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (404) 873-8794  
TELEFAX: (404) 873-8795  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-566-190-7

Query Match 92.1%; Score 35; DB 3; Length 7;  
Best Local Similarity 85.7%; Pred. No. 2e+05;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 YSPTSPS 7  
Db 1 YSPTSPT 7

RESULT 12  
US-08-566-190-16  
Sequence 16, Application US/08566190  
Patent No. 6090784  
GENERAL INFORMATION:  
APPLICANT: Warren, Stephen L.  
TITLE OF INVENTION: RNA Polymerase II Peptides and Methods of Use  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Patrea L. Pabst  
STREET: 2800 One Atlantic Center, 1201 West Peachtree Street  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: USA  
ZIP: 30309-3450  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/566,190  
FILING DATE: 1-DEC-1995  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Pabst, Patrea L.  
REGISTRATION NUMBER: 31,284  
REFERENCE/DOCKET NUMBER: YU116  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (404) 873-8794  
TELEFAX: (404) 873-8795  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-566-190-16

Query Match 92.1%; Score 35; DB 3; Length 7;  
Best Local Similarity 85.7%; Pred. No. 2e+05;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 YSPTSPS 7  
Db 1 YTPSPS 7

RESULT 13  
US-08-218-027A-2  
Sequence 2, Application US/08218027A  
Patent No. 5565327  
GENERAL INFORMATION:  
APPLICANT: Greenleaf, Arno L.  
APPLICANT: Lee, Jae M.  
APPLICANT: Hardin, Steven H.  
TITLE OF INVENTION: METHODS OF DIAGNOSING PARASITIC  
INFECTIONS AND OF TESTING DRUG SUSCEPTIBILITY OF PARASITES  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Kenneth D. Sibley  
STREET: P.O. Drawer 34009  
CITY: Charlotte  
STATE: No. 5565327th Carolina  
COUNTRY: USA  
ZIP: 28234  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/218,027A  
FILING DATE: 25-MAR-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Sibley, Kenneth D.  
REGISTRATION NUMBER: 31,665  
REFERENCE/DOCKET NUMBER: 5405-91  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-420-2200  
TELEFAX: 919-881-3175  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-218-027A-2

Query Match 89.5%; Score 34; DB 1; Length 7;  
Best Local Similarity 100.0%; Pred. No. 2e+05;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YSPTSP 6  
|||||  
DB 1 YSPTSP 6

RESULT 14  
US-08-218-027A-6  
Sequence 6, Application US/08218027A  
Patent No. 5565327  
GENERAL INFORMATION:  
APPLICANT: Greenleaf, Arno L.  
APPLICANT: Lee, Jae M.  
TITLE OF INVENTION: METHODS OF DIAGNOSING PARASITIC  
TITLE OF INVENTION: INFECTIONS AND OF TESTING DRUG SUSCEPTIBILITY OF PARASITES  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Kenneth D. Sibley  
STREET: P.O. Drawer 34009  
CITY: Charlotte  
STATE: No. 5565327th Carolina  
COUNTRY: USA  
ZIP: 28234  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/218,027A  
FILING DATE: 25-MAR-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Sibley, Kenneth D.  
REGISTRATION NUMBER: 31,665  
REFERENCE/DOCKET NUMBER: 5405-91  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-420-2200  
TELEFAX: 919-881-3175  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-218-027A-6

Query Match 89.5%; Score 34; DB 1; Length 7;  
Best Local Similarity 100.0%; Pred. No. 2e+05;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YSPTSP 6  
|||||  
DB 1 YSPTSP 6

RESULT 15  
US-08-348-718-3  
Sequence 3, Application US/08348718  
Patent No. 6007985  
GENERAL INFORMATION:  
APPLICANT: Stephen L. Warren  
TITLE OF INVENTION: RNA-Polymerase II Antibodies and  
TITLE OF INVENTION: Macromolecule Delivery Method  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Patrea L. Pabst  
STREET: 2800 One Atlantic Center  
STREET: 1201 West Peachtree Street  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: USA  
ZIP: 30309-3450  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/348,718  
FILING DATE: December 2, 1994  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/096,991  
FILING DATE: July 8, 1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/955,338  
FILING DATE: October 1, 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Pabst, Patrea L.  
REGISTRATION NUMBER: 31,284  
REFERENCE/DOCKET NUMBER: YU106CIP(2)  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (404) 873-8794  
TELEFAX: (404) 873-8795  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-348-718-3

Query Match 89.5%; Score 34; DB 3; Length 7;  
Best Local Similarity 100.0%; Pred. No. 2e+05;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YSPTSP 6  
|||||  
DB 1 YSPTSP 6

Search completed: June 23, 2003, 16:18:56  
Job time : 14 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2003, 16:17:26 ; Search time 19 seconds  
(without alignments)  
39.866 Million cell updates/sec

Title: US-09-733-773B-3

Perfect score: 38  
Sequence: 1 YSPSPS 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 417779 seqs, 108206813 residues

Total number of hits satisfying chosen parameters: 417779

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications.AA.\*  
1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PC1\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/PC07\_PUBCOMB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	38	100.0	314	9 US-10-043-487-311	Sequence 311, App
2	32	84.2	397	9 US-10-044-442-17	Sequence 17, Appl
3	32	84.2	2654	9 US-10-227-610-2	Sequence 2, Appl
4	31	81.6	37	10 US-09-864-761-36320	Sequence 36320, A
5	31	81.6	144	10 US-09-815-242-5486	Sequence 5486, Ap
6	31	81.6	149	10 US-09-815-242-12588	Sequence 12588, A
7	31	81.6	149	10 US-09-815-242-12936	Sequence 12936, A
8	31	81.6	149	10 US-09-815-242-13071	Sequence 13071, A
9	31	81.6	219	9 US-10-036-542-73	Sequence 73, Appl
10	31	81.6	288	9 US-10-036-542-98	Sequence 98, Appl
11	31	81.6	288	10 US-09-925-300-1208	Sequence 1208, Ap
12	31	81.6	327	9 US-09-573-446-3	Sequence 3, Appl
13	31	81.6	721	9 US-09-764-868-731	Sequence 731, Appl
14	31	81.6	2150	9 US-10-135-322-17	Sequence 17, Appl
15	30	78.9	318	10 US-09-867-550-1188	Sequence 1188, Ap
16	30	78.9	422	10 US-09-764-864-842	Sequence 842, App
17	30	78.9	422	10 US-09-764-864-1297	Sequence 1297, Ap
18	30	78.9	643	9 US-10-175-002-2	Sequence 2, Appl
19	30	78.9	1381	9 US-10-233-131-25	Sequence 25, Appl

20	29	76.3	43	9 US-09-989-920-266	Sequence 266, App
21	29	76.3	152	9 US-10-290-058A-15	Sequence 15, Appl
22	29	76.3	171	9 US-09-852-455-6	Sequence 6, Appl
23	29	76.3	202	9 US-09-738-626-3993	Sequence 3993, Ap
24	29	76.3	205	9 US-10-136-511-4	Sequence 4, Appl
25	29	76.3	205	10 US-09-879-919-6	Sequence 6, Appl
26	29	76.3	205	10 US-09-027-1287-4	Sequence 4, Appl
27	29	76.3	205	10 US-09-246-1298-4	Sequence 4, Appl
28	29	76.3	205	10 US-09-345-790-4	Sequence 4, Appl
29	29	76.3	205	10 US-09-899-059-4	Sequence 6, Appl
30	29	76.3	205	12 US-10-082-260-6	Sequence 20, Appl
31	29	76.3	381	9 US-10-165-603-20	Sequence 4, Appl
32	29	76.3	382	8 US-08-875-849C-4	Sequence 4, Appl
33	29	76.3	382	9 US-09-862-989-4	Sequence 70, Appl
34	29	76.3	406	10 US-09-859-214-70	Sequence 2, Appl
35	29	76.3	406	8 US-08-875-849C-2	Sequence 68, Appl
36	29	76.3	406	9 US-10-047-542-68	Sequence 2, Appl
37	29	76.3	406	9 US-09-862-989-2	Sequence 2, Appl
38	29	76.3	406	10 US-09-859-214-68	Sequence 68, Appl
39	29	76.3	510	10 US-09-866-562-56	Sequence 37, Appl
40	29	76.3	576	9 US-09-364-847-37	Sequence 49, Appl
41	29	76.3	712	9 US-09-364-847-49	Sequence 51, Appl
42	29	76.3	712	9 US-09-364-847-51	Sequence 36688, A
43	28	73.7	54	10 US-09-864-761-36688	Sequence 176, App
44	28	73.7	59	10 US-09-739-907-176	Sequence 96, Appl
45	28	73.7	60	10 US-09-739-907-96	

## ALIGNMENTS

RESULT 1  
US-10-043-487-311

Sequence 311, Application US/10043487  
Publication No. US20030055220A1

GENERAL INFORMATION:  
APPLICANT: HYBRIGENICS

TITLE OF INVENTION: Protein-Protein interactions between Shigella flexneri polypep

TITLE OF INVENTION: mammalian polypeptides

FILE REFERENCE: B4778A

CURRENT APPLICATION NUMBER: US/10/043,487

PRIOR FILING DATE: 2002-04-30

PRIOR APPLICATION NUMBER: US 60/261,130

PRIOR FILING DATE: 2001-01-12

NUMBER OF SEQ ID NOS: 561

SOFTWARE: PatentIn version 3.1

SEQ ID NO 311

LENGTH: 314

TYPE: PRT

ORGANISM: Shigella flexneri

US-10-043-487-311

Query Match

Best Local Similarity 100.0%; Score 38; DB 9; Length 314;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1 YSPSPS 7

RESULT 2

US-10-044-442-17

Sequence 17, Application US/10044442

Patent No. US20020160355A1

GENERAL INFORMATION:  
APPLICANT: Malcojm Whitman and Xin Chen

TITLE OF INVENTION: METHODS AND REAGENTS FOR MODULATING

TGF-BETA SUPERFAMILY SIGNALING

NUMBER OF SEQUENCES: 18

CORRESPONDENCE ADDRESS:

ADDRESSEE: Clark & Elding LLP

STREET: 176 Federal Street  
CITY: Boston  
STATE: MA  
COUNTRY: USA  
ZIP: 02110  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSO for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/044,442  
FILING DATE: 11-Jan-2002  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/087,134  
FILING DATE: 27-May-1998  
APPLICATION NUMBER: 60/047,991  
FILING DATE: 28-May-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Bleker-Brady, Kristina  
REGISTRATION NUMBER: 39,109  
REFERENCE/DOCKET NUMBER: 00246/501002  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-428-0200  
TELEFAX: 617-428-7045  
TELEX: <Unknown>  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 397 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FEATURE:  
NAME/KEY: Other  
LOCATION: 1...1  
OTHER INFORMATION: Mouse FAST-1  
SEQUENCE DESCRIPTION: SEQ ID NO: 17:  
US-10-044-442-17  
Query Match 84.2%; Score 32; DB 9; Length 397;  
Best Local Similarity 71.4%; Pred. No. 4.8e+03;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
QY 1 YSPSPS 7  
11 YTPPTPS 17  
Db  
RESULT 3  
US-10-227-610-2  
Sequence 2, Application US/10227610  
Publication No. US20030108916A1  
GENERAL INFORMATION:  
APPLICANT: Dunphy, W.  
APPLICANT: Kumagai, A.  
APPLICANT: Guo, Z.  
TITLE OF INVENTION: CLONING AND FUNCTIONAL ASSAYS OF XENOPUS ATR  
FILE REFERENCE: CICH-P01-003  
CURRENT APPLICATION NUMBER: US/10/227,610  
PRIOR FILING DATE: 2002-08-22  
PRIOR APPLICATION NUMBER: 60/314,215  
NUMBER OF SEQ ID NOS: 5  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 2  
LENGTH: 2654  
TYPE: PRT  
ORGANISM: Xenopus laevis  
US-10-227-610-2  
Query Match 84.2%; Score 32; DB 9; Length 2654;

Best Local Similarity 71.4%; Pred. No. 3.3e+03;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
QY 1 YSPSPS 7  
11 YTPPTPS 17  
Db 743 YSPAPT 749  
RESULT 4  
US-09-864-761-36320  
Sequence 36320, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chan, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO  
FILE REFERENCE: Acomlca-x-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
PRIOR FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 36320  
LENGTH: 37  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO ALO31983.1  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2  
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.5  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.6  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.4  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.6  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.3  
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.4



OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.3  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.4  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.5  
OTHER INFORMATION: EST\_HUMAN HIT: AM967376.1, EVALU6 6.50e+00  
US-09-864-761-36320

Query Match 81.6%; Score 31; DB 10; Length 37;  
Best Local Similarity 100.0%; Pred. No. 66;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 SPTSPTS 7  
|:|:|:|  
Db 26 SPTSPTS 31

RESULT 5  
US-09-815-242-5486  
Sequence 5486, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Karl L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 5486  
LENGTH: 144  
TYPE: PRT  
ORGANISM: Staphylococcus aureus  
US-09-815-242-5486

Query Match 81.6%; Score 31; DB 10; Length 144;  
Best Local Similarity 71.4%; Pred. No. 2; e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 YSPTSPTS 7  
|:|:|:|  
Db 98 YAPSSPS 104

RESULT 6  
US-09-815-242-12588  
Sequence 12588, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Karl L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel

APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 12588  
LENGTH: 149  
TYPE: PRT  
ORGANISM: Staphylococcus aureus  
US-09-815-242-12588

Query Match 81.6%; Score 31; DB 10; Length 149;  
Best Local Similarity 71.4%; Pred. No. 2; e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 YSPTSPTS 7  
|:|:|:|  
Db 102 YAPSSPS 108

RESULT 7  
US-09-815-242-12936  
Sequence 12936, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Karl L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE REFERENCE: ELITRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110

SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 12936  
LENGTH: 149  
TYPE: PRT  
ORGANISM: Staphylococcus aureus  
US-09-815-242-12936

Query Match  
Best Local Similarity 81.6%; Score 31; DB 10; Length 149;  
71.4%; Pred. No. 2.7e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 YSPSPS 7  
1111111  
Db 102 YAPSSPS 108

RESULT 8  
US-09-815-242-13071  
Sequence 13071, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Kari L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE REFERENCE: ELITRA.011A  
CURRENT FILING DATE: 2001-03-21  
PRIOR FILING DATE: 2000-03-21  
PRIOR FILING DATE: 2000-03-21  
PRIOR FILING DATE: 2000-05-23  
PRIOR FILING DATE: 2000-05-23  
PRIOR FILING DATE: 2000-05-26  
PRIOR FILING DATE: 2000-05-26  
PRIOR FILING DATE: 2000-05-26  
PRIOR FILING DATE: 2000-10-23  
PRIOR FILING DATE: 2000-10-23  
PRIOR FILING DATE: 2000-11-27  
PRIOR FILING DATE: 2000-11-27  
PRIOR FILING DATE: 2000-12-22  
PRIOR FILING DATE: 2000-12-22  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 13071  
LENGTH: 149  
TYPE: PRT  
ORGANISM: Staphylococcus aureus  
US-09-815-242-13071

Query Match  
Best Local Similarity 81.6%; Score 31; DB 10; Length 149;  
71.4%; Pred. No. 2.7e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 YSPSPS 7  
1111111  
Db 102 YAPSSPS 108

RESULT 9  
US-10-036-542-73  
Sequence 73, Application US/10036542  
Publication No. US20030083481A1  
GENERAL INFORMATION:  
APPLICANT: Birse et al.  
TITLE OF INVENTION: 25 Human Prostate and Prostate Cancer Associated Proteins  
FILE REFERENCE: PA002P1

CURRENT APPLICATION NUMBER: US/10/036,542  
CURRENT FILING DATE: 2002-01-07  
PRIOR APPLICATION NUMBER: PCT/US00/19666  
PRIOR FILING DATE: 2000-07-20  
PRIOR APPLICATION NUMBER: 60/144,972  
PRIOR FILING DATE: 1999-07-21  
PRIOR APPLICATION NUMBER: 60/148,681  
PRIOR FILING DATE: 1999-08-13  
PRIOR APPLICATION NUMBER: 60/149,173  
PRIOR FILING DATE: 1999-08-17  
PRIOR FILING DATE: 1999-10-06  
PRIOR FILING DATE: 1999-10-06  
PRIOR APPLICATION NUMBER: 60/158,004  
PRIOR FILING DATE: 2000-04-05  
NUMBER OF SEQ ID NOS: 157  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 73  
LENGTH: 219  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-036-542-73

Query Match  
Best Local Similarity 81.6%; Score 31; DB 9; Length 219;  
71.4%; Pred. No. 3.9e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 YSPSPS 7  
1111111  
Db 129 YSPTEPN 135

RESULT 10  
US-10-036-542-98  
Sequence 98, Application US/10036542  
Publication No. US20030083481A1  
GENERAL INFORMATION:  
APPLICANT: Birse et al.  
TITLE OF INVENTION: 25 Human Prostate and Prostate Cancer Associated Proteins  
FILE REFERENCE: PA002P1  
CURRENT APPLICATION NUMBER: US/10/036,542  
CURRENT FILING DATE: 2002-01-07  
PRIOR APPLICATION NUMBER: PCT/US00/19666  
PRIOR FILING DATE: 2000-07-20  
PRIOR FILING DATE: 2000-07-20  
PRIOR FILING DATE: 1999-07-21  
PRIOR FILING DATE: 1999-07-21  
PRIOR FILING DATE: 1999-08-13  
PRIOR FILING DATE: 1999-08-13  
PRIOR FILING DATE: 1999-08-17  
PRIOR FILING DATE: 1999-08-17  
PRIOR FILING DATE: 1999-10-06  
PRIOR FILING DATE: 1999-10-06  
PRIOR APPLICATION NUMBER: 60/158,004  
PRIOR FILING DATE: 2000-04-05  
NUMBER OF SEQ ID NOS: 157  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 98  
LENGTH: 288  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (277)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-10-036-542-98

Query Match  
Best Local Similarity 81.6%; Score 31; DB 9; Length 288;  
71.4%; Pred. No. 5.2e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 YSPSPS 7  
1111111  
Db 247 YSPTEPN 253

RESULT 11  
US-09-925-300-1208  
; Sequence 1208, Application US/09925300  
; Patent No. US20020151681A1  
; GENERAL INFORMATION:  
; APPLICANT: Craig Rosen,  
; APPLICANT: Steve Ruben,  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA101  
; CURRENT APPLICATION NUMBER: US/09/925,300  
; PRIOR FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05988  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 1890  
; SOFTWARE: Patentln Ver. 2.0  
; SEQ ID NO 1208  
; LENGTH: 288  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (277)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-925-300-1208

Query Match  
Best Local Similarity 81.6%; Score 31; DB 10; Length 288;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 YSPSPS 7  
|||||  
Db 247 YSPTEPN 253

RESULT 12  
US-09-573-446-3  
; Sequence 3, Application US/09573446  
; Publication No. US20030013853A1  
; GENERAL INFORMATION:  
; APPLICANT: Bandman, Olga  
; APPLICANT: Shah, Purvi  
; APPLICANT: Guegler, Karl J.  
; APPLICANT: Corley, Neil C.  
; TITLE OF INVENTION: HUMAN AFLATOXIN B1 ALDEHYDE REDUCTASE  
; NUMBER OF SEQUENCES: 3  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/573,446  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/907,674  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0362 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555

TELEFAX: 415-845-4166  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 327 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: GenBank  
; CLONE: 433611  
US-09-573-446-3

Query Match  
Best Local Similarity 81.6%; Score 31; DB 9; Length 327;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 YSPSPS 7  
|||||  
Db 252 YGTPAPS 258

RESULT 13  
US-09-764-868-731  
; Sequence 731, Application US/09764868  
; Patent No. US20020168711A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PT32  
; CURRENT APPLICATION NUMBER: US/09/764,868  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - refer to PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 1510  
; SOFTWARE: Patentln Ver. 2.0  
; SEQ ID NO 731  
; LENGTH: 721  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (123)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-764-868-731

Query Match  
Best Local Similarity 81.6%; Score 31; DB 9; Length 721;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 SPSPS 7  
|||||  
Db 502 SPSPS 507

RESULT 14  
US-10-135-322-17  
; Sequence 17, Application US/10135322  
; Patent No. US20020173017A1  
; GENERAL INFORMATION:  
; APPLICANT: BENFEY, PN  
; APPLICANT: HELARIUTTA, Y  
; APPLICANT: MAHONEN, AP  
; APPLICANT: BONKE, AMM  
; APPLICANT: KAUPPINEN, L  
; APPLICANT: RITIKONEN, M  
; TITLE OF INVENTION: WOODEN LEG GENE, PROMOTER AND USES THEREOF  
; FILE REFERENCE: 5914-086-999  
; CURRENT APPLICATION NUMBER: US/10/135,322  
; CURRENT FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: 60/253,739  
; PRIOR FILING DATE: 2000-11-29  
; NUMBER OF SEQ ID NOS: 43  
; SOFTWARE: Patentln version 3.0

Mon Jun 23 16:45:03 2003

SEQ ID NO 17  
 LENGTH: 2150  
 TYPE: PRT  
 ORGANISM: Arabidopsis thaliana  
 US-10-135-322-17

Query Match 81.6%; Score 31; DB 9; Length 2150;  
 Best Local Similarity 100.0%; Pred. No. 3.9e+03;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 SPTSPTS 7  
 |||||  
 Db 96 SPTSPTS 101

RESULT 15  
 US-09-867-550-1188  
 Sequence 1188, Application US/09867550  
 Patent No. US20020082206A1  
 GENERAL INFORMATION:  
 APPLICANT: Leach, Martin D.  
 APPLICANT: Mehraban, Fuad,  
 APPLICANT: Conley, Pamela  
 APPLICANT: Law, Debbie  
 APPLICANT: Topper, James  
 TITLE OF INVENTION: No. US20020082206A1 Polynucleotides from Atherogenic Cells and  
 TITLE OF INVENTION: Thereby  
 FILE REFERENCE: 21402-013 (Cura-313)  
 CURRENT APPLICATION NUMBER: US/09/867,550  
 CURRENT FILING DATE: 2001-09-20  
 PRIOR APPLICATION NUMBER: USSN 60/208,427  
 PRIOR FILING DATE: 2000-05-30  
 NUMBER OF SEQ ID NOS: 2125  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 1188  
 LENGTH: 318  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 US-09-867-550-1188

Query Match 78.9%; Score 30; DB 10; Length 318;  
 Best Local Similarity 83.3%; Pred. No. 8.5e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 YSPISP 6  
 :|||||  
 Db 254 FSPISP 259

Search completed: June 23, 2003, 16:19:23  
 Job time : 21 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: June 23, 2003, 16:16:01; Search time 15 Seconds  
(without alignments)  
44.863 Million cell updates/sec

Title: US-09-733-773B-3

Sequence: 1 YSP7SPS 7

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

## Database:

PIR-73:\*  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38	100.0	467	2	A27677 DNA-directed RNA p
2	38	100.0	491	2	S14182 DNA-directed RNA p
3	38	100.0	625	2	T07796 DNA-directed RNA p
4	38	100.0	650	2	S14181 DNA-directed RNA p
5	38	100.0	902	2	A56823 DNA-directed RNA p
6	38	100.0	977	2	S14183 DNA-directed RNA p
7	38	100.0	1605	2	T31435 DNA-directed RNA p
8	38	100.0	1685	2	T43217 DNA polymerase (EC
9	38	100.0	1733	1	RNBY2L DNA-directed RNA p
10	38	100.0	1752	1	S26849 DNA-directed RNA p
11	38	100.0	1834	1	JDM01 DNA-directed RNA p
12	38	100.0	1840	2	G85422 hypothetical prote
13	38	100.0	1859	1	A34092 DNA-directed RNA p
14	38	100.0	1862	2	T29959 DNA-directed RNA p
15	38	100.0	1896	1	RNFR2L DNA-directed RNA p
16	38	100.0	1932	1	A28490 DNA-directed RNA p
17	38	100.0	1970	1	S21054 DNA-directed RNA p
18	38	100.0	1970	2	T38186 RNA polymerase II
19	35	92.1	204	2	PS0262 DNA-directed RNA p
20	35	92.1	455	2	A87343 glycosyl transfera
21	35	92.1	1553	2	T31436 DNA-directed RNA p
22	35	92.1	2452	1	RN202L DNA-directed RNA p
23	35	92.1	2457	2	T18492 hypothetical prote
24	34	89.5	192	2	T15764 hypothetical prote
25	34	89.5	405	1	WMNVPI immediate-early pr
26	34	89.5	405	2	T10420 immediate early pr
27	34	89.5	619	2	T26453 hypothetical prote
28	34	89.5	628	2	T39494 kinesin-like prote
29	34	89.5	1732	2	T43026 probable DNA-direc

30	34	89.5	3005	2	S33642 homeotic protein z
31	32	84.2	53	2	C82708 hypothetical prote
32	32	84.2	168	2	S69314 probable membrane
33	32	84.2	351	1	VCVETC coat protein - tur
34	32	84.2	353	2	G91169 probable O-methyl
35	32	84.2	428	2	T05253 CER2 protein homol
36	32	84.2	2338	2	T73957 kinase-related pro
37	32	84.2	2340	2	I48310 kinase-related pro
38	32	84.2	2347	1	TVHURS kinase-related pro
39	31	81.6	111	2	S54635 probable membrane
40	31	81.6	145	2	G83994 hypothetical prote
41	31	81.6	174	2	B89931 conserved hypothet
42	31	81.6	231	2	F84426 hypothetical prote
43	31	81.6	241	2	S24544 nitrate reductase
44	31	81.6	269	2	H81278 phosphatidate cyti
45	31	81.6	269	2	JS0171 chlorophyll a/b-bi

## ALIGNMENTS

RESULT 1  
A27677  
DNA-directed RNA polymerase (EC 2.7.7.6) II largest chain - Chinese hamster (fragment  
C:Species: Cricetus griseus (Chinese hamster)  
C>Date: 31-Mar-1989 #sequence, revision 31-Mar-1989 #text\_change 18-Jun-1999  
C:Accession: A27677  
R:Allison, L.A.; Wong, J.K.C.; Fitzpatrick, V.D.; Moye, M.; Ingles, C.J.  
Mol. Cell. Biol. 8, 321-329, 1988  
A:Title: The C-terminal domain of the largest subunit of RNA polymerase II of Sacchar  
A:Reference number: A93104; MUID:88094402; PMID:3122024  
A:Accession: A27677  
A:Molecule type: DNA  
A:Residues: 1-467 <ALL>  
A:Cross-references: GB:M19538; MID:9191170; PIDN:AAA37008.1; PID:9387055  
C:Genetics:  
A:Gene: RPO21  
A:introns: 33/1: 83/1  
C:Superfamily: human DNA-directed RNA polymerase II largest chain  
C:Keywords: DNA binding; nucleotidyltransferase; tandem repeat; transcription; zinc f

Query Match 100.0%; Score 38; DB 2; Length 467;  
Best local Similarity 100.0%; Pred. No. 14;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YSP7SPS 7  
Db 112 YSP7SPS 118

RESULT 2  
S14182  
DNA-directed RNA polymerase (EC 2.7.7.6) largest chain (isoform B2) - soybean (fragme  
C:Species: Glycine max (soybean)  
C>Date: 12-Feb-1993 #sequence, revision 12-Feb-1993 #text\_change 12-Sep-1997  
C:Accession: S14182  
R:Dietrich, M.A.; Prenger, J.P.; Gullfoyle, T.J.  
Plant Mol. Biol. 15, 207-223, 1990  
A:Title: Analysis of the genes encoding the largest subunit of RNA polymerase II in A  
A:Reference number: S11960; MUID:9135869; PMID:2103447  
A:Accession: S14182  
A:Molecule type: mRNA  
A:Residues: 1-491 <DIE>  
A:Cross-references: EMBL:X52493  
A:Note: translation of the nucleotide sequence is not complete  
C:Genetics:  
A:Gene: RPB1-B2  
A:introns: 416/2; 440/2; 458/2  
C:Superfamily: human DNA-directed RNA polymerase II largest chain  
C:Keywords: DNA binding; nucleotidyltransferase; tandem repeat; transcription; zinc f  
F:201-474/Region: 7-residue repeats

Query Match 100.0%; Score 38; DB 2; Length 491;

Best Local Similarity 100.0%; Pred. No. 15;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 YSPTSPS 7  
|||||||  
Db 263 YSPTSPS 269

## RESULT 3

T07796

DNA-directed RNA polymerase (EC 2.7.7.6) largest chain - soybean (fragment)

C:Species: Glycine max (soybean)

C>Date: 14-May-1999 #sequence\_revision 14-May-1999 #text\_change 08-Oct-1999

C:Accession: T07796

R:Detrich, M.A.; Prenger, J.P.; Gullfoyle, T.J.

Plant Mol. Biol. 15, 207-223, 1990

A:Title: Analysis of the genes encoding the largest subunit of RNA polymerase II in Arab

A:Reference number: S11960; MUID:91355869; PMID:2103447

A:Accession: T07796

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-625 <DIE>

A:Cross-references: EMBL:X52492; NID:g18731; PIDN:CAA36733.1; PID:g18732

A:Experimental source: strain var. Wayne

C:Genetics:

A:Gene: RPBI-B1

A>Note: Intron positions not resolved (incomplete sequence)

C:Superfamily: human DNA-directed RNA polymerase II largest chain

C:Keywords: DNA binding; nucleotidyltransferase; transcription

## Query Match

Best Local Similarity 100.0%; Score 38; DB 2; Length 625;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 YSPTSPS 7  
|||||||  
Db 419 YSPTSPS 425

## RESULT 4

S14181

DNA-directed RNA polymerase (EC 2.7.7.6) largest chain (isoform B1) - soybean (fragment)

C:Species: Glycine max (soybean)

C>Date: 12-Feb-1993 #sequence\_revision 12-Feb-1993 #text\_change 12-Sep-1997

C:Accession: S14181

R:Detrich, M.A.; Prenger, J.P.; Gullfoyle, T.J.

Plant Mol. Biol. 15, 207-223, 1990

A:Title: Analysis of the genes encoding the largest subunit of RNA polymerase II in Arab

A:Reference number: S11960; MUID:91355869; PMID:2103447

A:Accession: S14181

A:Molecule type: DNA

A:Residues: 1-650 <DIE>

A:Cross-references: EMBL:X52492

A>Note: translation of the nucleotide sequence is not complete

C:Genetics:

A:Gene: RPBI-B1

A:introns: 575/2; 599/2; 617/2

C:Superfamily: human DNA-directed RNA polymerase II largest chain

C:Keywords: DNA binding; nucleotidyltransferase; tandem repeat; transcription; zinc fing

F;360-633/Region: 7-residue repeats

## Query Match

Best Local Similarity 100.0%; Score 38; DB 2; Length 650;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 YSPTSPS 7  
|||||||  
Db 415 YSPTSPS 421

## RESULT 5

A56823

DNA-directed RNA polymerase (EC 2.7.7.6) II largest chain - slime mold (Dictyostelium d)

C:Species: Dictyostelium discoideum

C>Date: 05-Jan-1996 #sequence\_revision 05-Jan-1996 #text\_change 18-Jun-1999

C:Accession: A56823

R:Lam, T.Y.; Chan, L.; Yip, P.; Siu, C.H.

Biochem. Cell Biol. 70, 792-799, 1992

A:Title: The largest subunit of RNA polymerase II in Dictyostelium: conservation of t

A:Reference number: A56823; MUID:93129425; PMID:1482555

A:Accession: A56823

A:Molecule type: mRNA

A>Status: preliminary

A:Residues: 1-902 <LAM>

A:Cross-references: GB:S52651; NID:g263753; PIDN:AAB24966.1; PID:g263754

A>Note: sequence extracted from NCBI backbone (NCBIN:129221, NCBIPI:129222)

C:Superfamily: human DNA-directed RNA polymerase II largest chain

C:Keywords: DNA binding; nucleotidyltransferase; tandem repeat; transcription; zinc f

## Query Match

Best Local Similarity 100.0%; Score 38; DB 2; Length 902;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 YSPTSPS 7  
|||||||  
Db 728 YSPTSPS 734

## RESULT 6

S14183

DNA-directed RNA polymerase (EC 2.7.7.6) largest chain (isoform C) - soybean (fragmen

C:Species: Glycine max (soybean)

C>Date: 12-Feb-1993 #sequence\_revision 12-Feb-1993 #text\_change 18-Jun-1999

C:Accession: S14183

R:Detrich, M.A.; Prenger, J.P.; Gullfoyle, T.J.

Plant Mol. Biol. 15, 207-223, 1990

A:Title: Analysis of the genes encoding the largest subunit of RNA polymerase II in A

A:Reference number: S11960; MUID:91355869; PMID:2103447

A:Accession: S14183

A:Molecule type: DNA

A:Residues: 1-977 <DIE>

A:Cross-references: EMBL:X52495; NID:g18735; PIDN:CAA36736.1; PID:g18736

A>Note: translation of the nucleotide sequence is not complete

C:Genetics:

A:Gene: RPBI-C

A:introns: 906/2; 930/2; 948/2

C:Superfamily: human DNA-directed RNA polymerase II largest chain

C:Keywords: DNA binding; nucleotidyltransferase; tandem repeat; transcription; zinc f

F;725-964/Region: 7-residue repeats

## Query Match

Best Local Similarity 100.0%; Score 38; DB 2; Length 977;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 YSPTSPS 7  
|||||||  
Db 780 YSPTSPS 786

## RESULT 7

T31435

DNA-directed RNA polymerase (EC 2.7.7.6) II largest chain - Valiriorpha necatrix

C:Species: Valiriorpha necatrix

C>Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 19-May-2000

C:Accession: T31435

R:Hirt, R.P.; Logsdon, J.M.; Healy, B.; Dorey, M.W.; Doolittle, W.F.; Embley, T.M.

Proc. Natl. Acad. Sci. U.S.A. 96, 580-585, 1999

A:Title: Microsporidia are related to fungi: evidence from the largest subunit of RNA

A:Reference number: Z21034; MUID:91110933; PMID:9682676

A:Accession: T31435

A:Molecule type: DNA

A>Status: preliminary

A:Residues: 1-1605 <HIR>

A:Cross-references: EMBL:AA060234; NID:g4001823; PID:g4001824; PIDN:AA012604.1

C:Superfamily: human DNA-directed RNA polymerase II largest chain

C:Keywords: nucleotidyltransferase

Query Match 100.0%; Score 38; DB 2; Length 1605;  
 Best Local Similarity 100.0%; Pred. No. 53;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YSPTSPS 7  
 |||||  
 DB 1509 YSPTSPS 1515

## RESULT 8

T43217

RNA polymerase (EC 2.7.7.-) II, largest chain - *Porphyra yezoensis* (fragment)  
 C:Species: *Porphyra yezoensis*  
 C:Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 21-Jul-2000  
 C:Accession: T43217  
 R:Stiller, J.W.; Hall, B.D.  
 Submitted to the EMBL Data Library, June 1998  
 A:Reference number: Z22292  
 A:Accession: T43217  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-1685 <STR>  
 A:Cross-references: EMBL:U90208; NID:g3169690; PID:g3169691; PIDN:AAC17924.1  
 A:Experimental source: strain U51  
 C:Genetics:  
 A:Gene: rpbl  
 C:Superfamily: human DNA-directed RNA polymerase II largest chain  
 C:Keywords: nucleotidyltransferase

Query Match 100.0%; Score 38; DB 2; Length 1685;  
 Best Local Similarity 100.0%; Pred. No. 56;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YSPTSPS 7  
 |||||  
 DB 1612 YSPTSPS 1618

## RESULT 9

RNB2L

DNA-directed RNA polymerase (EC 2.7.7.6) II 215K chain - yeast (*Saccharomyces cerevisiae*)  
 M:Alternate names: protein D2150; protein YDL140c; RNA polymerase II largest chain  
 C:Species: *Saccharomyces cerevisiae*  
 C:Date: 04-Dec-1986 #sequence\_revision 23-Aug-1996 #text\_change 21-Jul-2000  
 C:Accession: S67686; A00692; S64647; S20985  
 R:Saluz, H.P.; Woelfli, S.; Hanemann, V.  
 Submitted to the Protein Sequence Database, July 1996  
 A:Reference number: S67677  
 A:Accession: S67686  
 A:Molecule type: DNA  
 A:Residues: 1-1733 <SAL>  
 A:Cross-references: EMBL:Z74188; NID:g1431216; PIDN:CAA98713.1; PID:g1431217; GSPDB:GNOC  
 R:Allison, L.A.; Moyie, M.; Shales, M.; Ingles, C.J.  
 Cell 42, 599-610, 1985  
 A:Title: Extensive homology among the largest subunits of eukaryotic and prokaryotic RNA  
 A:Reference number: A90870; MUID:85282617; PMID:3896517  
 A:Accession: A00692  
 A:Molecule type: DNA

A:Residues: 1-1513, 'V', 1515-1523, 'A', 1525-1555, 1563-1607, 'M', 1609-1733 <ALL>  
 R:Cross-references: EMBL:X03128; NID:g43397; PIDN:CAA26904.1; PID:g43398  
 R:Cronan Jr., J.E.; Wallace, J.C.  
 FEMS Microbiol. Lett. 130, 221-230, 1995  
 A:Title: The gene encoding the biotin-*apoprotein* ligase of *Saccharomyces cerevisiae*.  
 A:Reference number: S64646; MUID:95377607; PMID:7649444  
 A:Accession: S64647  
 A:Status: translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1669-1733 <CRO>  
 A:Cross-references: EMBL:U27182; NID:g886080; PIDN:AAC49058.1; PID:g886082  
 C:Genetics:  
 A:Gene: SGD:RP021; MIPS:YDL140C

A:Cross-references: SGD:S0002299; MIPS:YDL140C  
 A:Map position: 4L  
 C:Superfamily: human DNA-directed RNA polymerase II largest chain  
 C:Keywords: DNA binding; nucleotidyltransferase; tandem repeat; transcription; zinc f  
 F:1567-1733/Region: 7-residue repeats

Query Match 100.0%; Score 38; DB 1; Length 1733;  
 Best Local Similarity 100.0%; Pred. No. 58;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YSPTSPS 7  
 |||||  
 DB 1556 YSPTSPS 1562

## RESULT 10

S26849

DNA-directed RNA polymerase (EC 2.7.7.6) II largest chain - fission yeast (*Schizosac*  
 C:Species: *Schizosaccharomyces pombe*  
 C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Dec-1999  
 C:Accession: S26849; T40055  
 R:Aizuma, Y.; Yamagishi, M.; Ueshima, R.; Ishihama, A.  
 Nucleic Acids Res. 19, 461-468, 1991  
 A:Title: Cloning and sequence determination of the *Schizosaccharomyces pombe* rpbl gen  
 A:Reference number: S26849; MUID:91187661; PMID:2011520  
 A:Accession: S26849

A:Molecule type: DNA  
 A:Residues: 1-1752 <AZU>  
 A:Cross-references: EMBL:X56564; NID:g50554; PIDN:CAA39916.1; PID:g5055  
 A:Note: the authors did not translate the codon for residue 1464  
 R:McDougal, R.C.; Rajandream, M.A.; Barrell, B.G.; Stevens, K.; Badcock, K.; Church  
 submitted to the EMBL Data Library, October 1999  
 A:Reference number: Z21902  
 A:Accession: T40055  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1752 <MCD>  
 A:Cross-references: EMBL:AL121795; PIDN:CAB57941.1; GSPDB:GN00067; SPDB:SPBC28F2.12  
 C:Genetics:  
 A:Gene: SPBC28F2.12  
 A:Map position: 2  
 A:Insertion: 5/3; 14/2; 38/3; 64/1; 84/1; 119/3  
 A:Superfamily: human DNA-directed RNA polymerase II largest chain  
 C:Keywords: DNA binding; nucleotidyltransferase; tandem repeat; transcription; zinc f

Query Match 100.0%; Score 38; DB 1; Length 1752;  
 Best Local Similarity 100.0%; Pred. No. 58;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YSPTSPS 7  
 |||||  
 DB 1578 YSPTSPS 1584

## RESULT 11

JDM11

DNA-directed RNA polymerase (EC 2.7.7.6) II largest chain - *Arabidopsis thaliana*  
 M:Alternate names: DNA-directed RNA polymerase II 205K chain; protein FAB14.70  
 C:Species: *Arabidopsis thaliana* (mouse-ear cress)  
 C:Date: 31-Mar-1993 #sequence\_revision 11-Jun-1999 #text\_change 20-Aug-1999  
 C:Accession: T04690; S12071; S27346; S11960  
 R:Bayan, M.; Rose, M.; Hempel, S.; Ertlan, K.D.; Hohenseil, J.; Mewes, H.W.; Mayer, K.  
 submitted to the Protein Sequence Database, October 1998  
 A:Reference number: Z15380  
 A:Accession: T04690  
 A:Molecule type: DNA  
 A:Residues: 1-1834 <BEV>  
 A:Cross-references: EMBL:AL031986  
 A:Experimental source: cultivar Columbia; BAC clone FAB14  
 R:Newirth, C.; Scheill, J.; Koncz, C.  
 Mol. Gen. Genet. 223, 65-75, 1990  
 A:Title: Homologous domains of the largest subunit of eucaryotic RNA polymerase II ar

A:Reference number: S12071; MUID:91080867; PMID:2259344  
 A:Accession: S12071  
 A:Molecule type: DNA  
 A:Residues: 1-421,'S',423-732,'D',734-1055,'R',1057-1714,'SPTSPPSY',1715-1834 <NMM1>  
 A:Cross-references: EMBL:X52954; NID:916504; PIDN:CAA37130.1; PID:916505  
 A:Experimental source: cv. Columbia  
 A>Note: the authors translated the codon AGC for residue 1755 as Arg  
 A:Accession: S27346  
 A:Molecule type: mRNA  
 A:Residues: 510-732,'D',734-1055,'R',1057-1714,'SPTSPPSY',1715-1834 <NMM2>  
 A:Experimental source: cv. Columbia  
 A:Dictator: M.A.; Prenger, J.P.; Guilfoyle, T.J.  
 A:Title: Biol. 15, 207-223, 1990  
 A:Title: Analysis of the genes encoding the largest subunit of RNA polymerase II in Arab  
 A:Accession: S11960  
 A:Residues: 1-116,125-193,'NSKEE',198-297,'R',299-302,'R',304-400,'KE',403,'VDYGPHPGK  
 A:Cross-references: EMBL:X52494; NID:916493; PIDN:CAA36735.1; PID:916494  
 A>Note: the authors translated the codon CCT for residue 1083 as Ala  
 C:Genetics:  
 A:Gene: p11215; RPB1  
 A:Map position: 4  
 A:Introns: 28/3; 123/3; 218/3; 271/3; 325/3; 411/1; 440/3; 552/3; 648/2; 734/3; 1755/2;  
 A>Note: F4B14.70  
 C:Superfamily: human DNA-directed RNA polymerase II largest chain  
 C:Keywords: DNA binding; nucleotidyltransferase; phosphoprotein; tandem repeat; transcri  
 F:66-109/Region: zinc finger CCCC motif  
 F:1531-1812/Region: 7-residue repeats

Query Match 100.0%; Score 38; DB 1; Length 1834;  
 Best Local Similarity 100.0%; Pred. No. 61;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YSPSPS 7  
 |||||  
 DB 1588 YSPSPS 1594

RESULT 12  
 G85422  
 hypothetical protein AT4g35800 [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 02-Mar-2001  
 C:Accession: G85422  
 R:Anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Sprin  
 Nature 402, 769-777, 1999  
 A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.  
 A:Reference number: A85001; MUID:20083488; PMID:10617198  
 A:Accession: G85422  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1840 <SPD>  
 A:Cross-references: GB:NC\_001268; NID:97270532; PIDN:CAB81489.1; GSPDB:GN00140  
 C:Genetics:  
 A:Gene: AT4g35800  
 A:Map position: 4  
 C:Superfamily: human DNA-directed RNA polymerase II largest chain

Query Match 100.0%; Score 38; DB 2; Length 1840;  
 Best Local Similarity 100.0%; Pred. No. 62;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YSPSPS 7  
 |||||  
 DB 1594 YSPSPS 1600

RESULT 13  
 A34092  
 DNA-directed RNA polymerase (EC 2.7.7.6) II large chain - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999

C:Accession: A34092  
 R:Bird, D.M.; Riddle, D.L.  
 Mol. Cell. Biol. 9, 4119-4130, 1989  
 A:Title: Molecular cloning and sequencing of ama-1, the gene encoding the largest sub  
 A:Reference number: A34092; MUID:9006416; PMID:2586513  
 A:Accession: A34092  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1859 <BR>  
 A:Cross-references: GB:M29235; NID:9156404; PIDN:AAA28126.1; PID:9156405  
 A>Note: the authors failed to translated the codon GAG for residue 917 as Glu, and CA  
 C:Superfamily: human DNA-directed RNA polymerase II largest chain  
 C:Keywords: DNA binding; nucleotidyltransferase; tandem repeat; transcription; zinc f

Query Match 100.0%; Score 38; DB 1; Length 1859;  
 Best Local Similarity 100.0%; Pred. No. 62;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YSPSPS 7  
 |||||  
 DB 1596 YSPSPS 1602

RESULT 14  
 T29959  
 DNA-directed RNA polymerase II large subunit ama-1 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 11-Jan-2000  
 C:Accession: T29959  
 R:Miller, N.; Bradshaw, H.  
 submitted to the EMBL Data Library, April 1996  
 A:Description: The sequence of C. elegans cosmid F36A4.  
 A:Reference number: 220713  
 A:Accession: T29959  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1862 <MIL>  
 A:Cross-references: EMBL:053333; PIDN:AAA96158.1; GSPDB:GN00022; CESP:ama-1  
 A:Experimental source: strain Bristol N2; clone F36A4  
 C:Genetics:  
 A:Gene: CESP:ama-1  
 A:Map position: 4  
 A:Introns: 28/3; 65/2; 175/1; 411/2; 513/2; 773/3; 921/3; 1168/3; 1698/2; 1767/3; 182  
 C:Superfamily: human DNA-directed RNA polymerase II largest chain

Query Match 100.0%; Score 38; DB 2; Length 1862;  
 Best Local Similarity 100.0%; Pred. No. 62;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YSPSPS 7  
 |||||  
 DB 1599 YSPSPS 1605

RESULT 15  
 RNF2L  
 DNA-directed RNA polymerase (EC 2.7.7.6) II 215K chain [validated] - fruit fly (Droso  
 C:Species: Drosophila melanogaster  
 C:Date: 04-Dec-1986 #sequence\_revision 31-Mar-1993 #text\_change 28-Jul-2000  
 C:Accession: S04457; A00693; B27677; S60151  
 R:Jokier, R.S.; Weeks, J.R.; Zehring, W.A.; Greenleaf, A.L.  
 Mol. Gen. Genet. 215, 266-275, 1989  
 A:Title: Analysis of the gene encoding the largest subunit of RNA polymerase II in Dr  
 A:Reference number: S04457; MUID:89218930; PMID:2496296  
 A:Accession: S04457  
 A:Molecule type: DNA  
 A:Residues: 1-1896 <JOK>  
 A:Cross-references: EMBL:M27431; NID:9158331; PIDN:AAA28868.1; PID:9158332  
 R:Biggs, J.; Searles, L.L.; Greenleaf, A.L.  
 Cell 42, 611-621, 1985  
 A:Title: Structure of the eukaryotic transcription apparatus: features of the gene fo  
 A:Reference number: A00693; MUID:85282618; PMID:2992806  
 A:Accession: A00693



A:Molecule type: DNA  
 A:Residues: 1-318, 'GYAKV', 325-449, 'G', 451-454, 'RCTT', 459-462, 'VTGESVASS' <BIG>  
 A:Cross-references: EMBL:M11798  
 A>Note: this sequence has been revised in reference S04457  
 R:Allison, L.A.; Wong, J.K.C.; Fitzpatrick, V.D.; Moyle, M.; Ingles, C.J.  
 Mol. Cell. Biol. 8, 321-329, 1988  
 A>Title: The C-terminal domain of the largest subunit of RNA polymerase II of Saccharomyces cerevisiae  
 A:Reference number: A93104; MUID:88094402; PMID:3122024  
 A:Accession: B27677  
 A:Molecule type: DNA  
 A:Residues: 1441-1484, 'T', 1527-1889 <ALL>  
 A:Cross-references: EMBL:M19537; NID:9158147; PIDN:AAA28827.1; PID:9158148  
 R:Peterson, G.; Song, D.; Huegler-Doerr, B.; Oldenburg, I.; Bautz, E.K.F.  
 Mol. Gen. Genet. 249, 425-431, 1995  
 A>Title: Mapping of linear epitopes recognized by monoclonal antibodies with gene-fragments  
 A:Reference number: S60151; MUID:96133682; PMID:8552047  
 A:Accession: S60151  
 A>Status: preliminary; not compared with conceptual translation  
 A:Molecule type: DNA  
 A:Residues: 778-827 <PET>  
 C:Genetics:  
 A:Gene: RPO21; RPII215  
 A:Cross-references: FlyBase:FBgn0003277  
 A:Map position: X 10C, X 35.7  
 A:Introns: 27/3; 775/3; 1526/1  
 C:Function:  
 A:Description: EC 2.7.7.6 [validated, MUID:88094402]; essential for proper initiation of transcription  
 C:Superfamily: human DNA-directed RNA polymerase II largest chain  
 C:Keywords: DNA binding; nucleotidyltransferase; tandem repeat; transcription; zinc finger  
 F:67-83/Region: zinc finger CCHH motif  
 F:1581-1883/Region: 7-residue repeats  
 F:349/Binding site: ATP/GTP (Lys) #status predicted

Query Match 100.0%; Score 38; DB 1; Length 1896;  
 Best Local Similarity 100.0%; Pred. No. 64;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YSPTSPS 7  
 DB 1687 YSPTSPS 1693

Search completed: June 23, 2003, 16:18:36  
 Job time : 16 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2003, 16:13:06 ; Search time 11 Seconds

(without alignments)  
26.394 Million cell updates/sec

Title: US-09-733-773B-3

Perfect score: 38  
Sequence: 1 YSPRSPS 7

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues  
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	38	100.0	467	1 RPB1_CRIGR	P11414 Crictulius
2	38	100.0	902	1 RPB1_DICDI	P35084 dictyostei1
3	38	100.0	1733	1 RPB1_YEAST	P04050 saccharomyc
4	38	100.0	1752	1 RPB1_SCHPO	P36594 schizosacch
5	38	100.0	1840	1 RPB1_ARATH	P18616 arabidopsis
6	38	100.0	1852	1 RPB1_CAEEL	P16356 caenorhabdi
7	38	100.0	1887	1 RPB1_DROME	P04052 drosophila
8	38	100.0	1970	1 RPB1_HUMAN	P24928 homo sapien
9	38	100.0	1970	1 RPB1_MOUSE	P08775 mus musculu
10	35	92.1	2452	1 RPB1_PLAFD	P14248 plasmodium
11	34	89.5	192	1 VIB7_CAEEL	P46581 caenorhabdi
12	34	89.5	405	1 VIE2_NPYOP	P32511 oryza pseu
13	34	89.5	3005	1 ZFH2_DROME	P28167 drosophila
14	34	84.2	351	1 COAT_TCV	P06663 turrip crin
15	32	84.2	401	1 FXH1_MOUSE	O88621 mus musculu
16	32	84.2	2347	1 KROS_HUMAN	P08922 homo sapien
17	31	81.6	231	1 NTA2_MATZE	P39871 zea mays (m
18	31	81.6	269	1 CB2_PHYRA	P20866 physcomitre
19	31	81.6	277	1 RM02_RECMA	O31247 recitmonoma
20	31	81.6	321	1 Y126_TREPA	O83163 treponema p
21	31	81.6	327	1 AR71_RAT	P38918 ratius norv
22	31	81.6	506	1 NFIC_PIG	P21999 sus scrofa
23	31	81.6	508	1 NFIC_HUMAN	P08651 homo sapien
24	31	81.6	515	1 HNSH_DROME	O03372 drosophila
25	31	81.6	595	1 YET2_HUMAN	O08435 homo sapien
26	31	81.6	744	1 IE63_HCVVA	P16749 human cylon
27	31	81.6	771	1 YKYO_YEAST	O02208 saccharomyc
28	31	81.6	801	1 DHGA_ACICA	P05465 acinetobact
29	31	81.6	982	1 POL_HTV2	P03363 human t-cel
30	31	81.6	992	1 DPLI_RAT	P97836 ratius norv
31	30	78.9	265	1 EN16_STRPU	P13665 strongyloce
32	30	78.9	329	1 PDXA_ECO57	P38713 escherichia
33	30	78.9	329	1 PDXA_ECOLI	P19624 escherichia

34	30	78.9	346	1 RPAF_HAEIN	P45042 haemophilus
35	30	78.9	349	1 STR2_STRGR	P29780 streptomyce
36	30	78.9	378	1 VGLM_HSYE2	P52371 equine herp
37	30	78.9	392	1 Y056_CAEEL	O09252 caenorhabdi
38	30	78.9	445	1 SLP2_DROME	P32031 drosophila
39	30	78.9	560	1 NF1B_CHICK	P17924 gallus gall
40	30	78.9	657	1 MYB_DROME	P04197 drosophila
41	30	78.9	796	1 DHG_ECOLI	P15877 escherichia
42	30	78.9	801	1 PIPA_DICDI	O02158 dictyostei1
43	30	78.9	808	1 GLT4_WHEAT	P27175 glucocobact
44	30	78.9	838	1 GLT4_WHEAT	P08489 triticum ae
45	30	78.9	839	1 GLT5_WHEAT	P10388 triticum ae

#### ALIGNMENTS

```

RESULT 1
RPB1_CRIGR
ID RPB1_CRIGR STANDARD: PRT: 467 AA.
AC P11414;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE DNA-directed RNA polymerase II largest subunit (EC 2.7.7.6) (RPB1)
DE (Fragment).
GN POLR2A.
OS Crictulius griseus (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Crictinae;
OC Crictulius.
OX NCBI_TaxID=10029;
RN [1]
RP MEDLINE=88094402; PubMed=3122024;
RA Allison L.A., Wong J.K.-C., Fitzpatrick V.D., Moyle M., Ingles C.J.;
RT "The C-terminal domain of the largest subunit of RNA polymerase II of
RT Saccharomyces cerevisiae, Drosophila melanogaster, and mammals: a
RT conserved structure with an essential function.";
RL Mol. Cell. Biol. 8:321-329(1988).
CC - FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC SUBSTRATES.
CC - CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC (RNA)(N).
CC - SUBUNIT: RNA POLYMERASE II CONSISTS OF 12 DIFFERENT SUBUNITS.
CC - SUBCELLULAR LOCATION: Nuclear.
CC - PTM: THE TANDEN 7 RESIDUES REPEATS CAN BE HIGHLY PHOSPHORYLATED.
CC THE PHOSPHORYLATION ACTIVATES POL2.
CC - MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE
CC FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA
CC PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE
CC III FOR 5S AND 5.8S GENES.
CC - SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.
CC
CC THIS SWISS-PROT ENTRY IS COPYRIGHTED. IT IS PRODUCED THROUGH A COLLABORATION
CC BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION -
CC THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS
CC USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY
CC MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL
CC ENTITIES REQUIRES A LICENSE AGREEMENT (See http://www.isb-sdb.ch/announce/
CC or send an email to license@isb-sdb.ch).
CC
CC EMBL: M19538; AAA37008.1; -.
CC PIR: A27677; A27677.
CC INTERPRO: IPR000684; RNA_POL11_repeat.
CC PROSITE: PS00115; RNA_POL_II_REPEAT; 43.
CC TRANSFERASE: DNA-directed RNA polymerase; Transcription; Zinc; Repeat;
CC DNA-binding; Nuclear protein; Phosphorylation.
CC NON_TER 1
CC DOMAIN 90 457 CARBOXYL-TERMINAL 7-RESIDUE REPEATS.
CC SEQUENCE 467 AA; 48167 MW; 0AADFD923B5D2BCE CRC64;

```

Query Match 100.0%; Score 38; DB 1; Length 467;  
 Best Local Similarity 100.0%; Pred. No. 6.1;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 YSPTSPS 7  
 DB 112 YSPTSPS 118

## RESULT 2

RPBL\_DICDI STANDARD; PRT; 902 AA.  
 AC P35064;  
 DT 01-FEB-1994 (Rel. 28, Created)  
 DT 01-FEB-1994 (Rel. 28, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE DNA-directed RNA polymerase II largest subunit (EC 2.7.7.6)  
 DE (Fragment).  
 GN RPOA.  
 OS Dictyostelium discoideum (Slime mold).  
 OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.  
 OX NCBI\_TaxID=44689;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RX MEDLINE=93129425; PubMed=1482555;  
 RA Lam T.Y., Chan L., Yip P., Siu C.H.;  
 "The largest subunit of RNA polymerase II in Dictyostelium:  
 RT conservation of the unique tail domain and gene expression.";  
 RT Biochem. Cell Biol. 70:792-799(1992).  
 CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION  
 CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS  
 CC SUBSTRATES.

CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +  
 CC [RNA](N).  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- PTM: THE TANDEM 7 RESIDUES REPEATS CAN BE HIGHLY PHOSPHORYLATED.  
 CC -1- THE PHOSPHORYLATION ACTIVATES POL2.  
 CC -1- MISCELLANEOUS: THREE DISTINCT RNA POLYMERASES WERE FOUND IN  
 CC EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA PRECURSOR,  
 CC POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE III FOR 5S  
 CC AND TRNA GENES.

CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

CC EMBL: S52651; AAB24966.1; -  
 DR Dictydb; DD01054; RPOA.  
 DR InterPro; IPR000684; RNA\_polII\_repeat.  
 DR InterPro; IPR000722; RNA\_pol\_A.  
 DR InterPro; IPR002879; RNA\_pol\_A2.  
 DR Pfam; PF00623; RNA\_pol\_A; 1.  
 DR Pfam; PF01854; RNA\_pol\_A2; 1.  
 DR PROSITE; PS00115; RNA\_POL\_II\_REPEAT; 22.  
 KM Transferase; DNA-directed RNA polymerase; Transcription; Zinc; Repeat;  
 KM DNA-binding; Nuclear protein; Phosphorylation; Zinc-finger.  
 FT NON\_TER  
 FT DOMAIN 1 888 CARBOXY-TERMINAL 7-RESIDUE REPEATS.  
 SQ SEQUENCE 902 AA; 100176 MW; 201D02470322203C CRC64;

Query Match 100.0%; Score 38; DB 1; Length 902;  
 Best Local Similarity 100.0%; Pred. No. 12;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 YSPTSPS 7  
 DB 728 YSPTSPS 734

RESULT 3  
 RPBL\_YEAST STANDARD; PRT; 1733 AA.

AC P04050; Q12364; Q92315;  
 DT 01-NOV-1986 (Rel. 03, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE DNA-directed RNA polymerase II largest subunit (EC 2.7.7.6) (R220).  
 GN RPB1 OR RPO21 OR RPB220 OR SUB8 OR YDI140C OR D2150.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OX NCBI\_TaxID=4932;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RX STRAIN=A364A;  
 RX MEDLINE=85282617; PubMed=3896517;  
 RA Allison L.A., Moyle M., Shales M., Ingles C.J.;  
 RT "Extensive homology among the largest subunits of eukaryotic and  
 RT prokaryotic RNA polymerases.";  
 RT Cell 42:599-610(1985).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=S288C / FY1679;  
 RX MEDLINE=97127826; PubMed=8972577;  
 RA Woelfl S., Haneman V., Saluz H.P.;  
 RT "Analysis of a 26,756 bp segment from the left arm of yeast  
 RT chromosome IV.";  
 RT Yeast 12:1549-1554(1996).  
 RN [3]  
 RP SEQUENCE OF 1669-1733 FROM N.A.  
 RX STRAIN=S288C;  
 RX MEDLINE=95377607; PubMed=7649444;  
 RA Cronan J.E. Jr., Wallace J.C.;  
 RT "The gene encoding the biotin apoprotein ligase of Saccharomyces  
 RT cerevisiae.";  
 RT FEMS Microbiol. Lett. 130:221-230(1995).

CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION  
 CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS  
 CC SUBSTRATES.  
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +  
 CC [RNA](N).  
 CC -1- SUBUNIT: RNA POLYMERASE II CONSISTS OF 12 DIFFERENT SUBUNITS.  
 CC THIS SUBUNIT IS THE LARGEST COMPONENT OF RNA POLYMERASE II.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- PTM: THE TANDEM 7 RESIDUES REPEATS CAN BE HIGHLY PHOSPHORYLATED.  
 CC -1- THE PHOSPHORYLATION ACTIVATES POL2.  
 CC -1- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE  
 CC FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA  
 CC PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE  
 CC III FOR 5S AND TRNA GENES.

CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

CC EMBL: X03128; CAA26904.1; -  
 DR EMBL: X96876; CAA65619.1; -  
 DR EMBL: Z74188; CAA98713.1; -  
 DR EMBL: U27182; AAC49058.1; -  
 DR PIR: A00692; RNB21.  
 DR SGD: S0002299; RPO21.  
 DR InterPro; IPR000684; RNA\_polII\_repeat.  
 DR InterPro; IPR000722; RNA\_pol\_A.  
 DR InterPro; IPR002879; RNA\_pol\_A2.  
 DR Pfam; PF00623; RNA\_pol\_A; 1.  
 DR Pfam; PF01854; RNA\_pol\_A2; 1.

DR PROSITE: PS00115; RNA\_POL\_II\_REPEAT; 22.  
 KW Transferring: DNA-directed RNA polymerase; Transcription; Zinc; Repeat;  
 KM DNA-binding: Nuclear protein; Phosphorylation; Zinc-finger.  
 FT ZN\_FING 67 83 C2H2-TYPE (POTENTIAL).  
 FT DOMAIN 1544 1719 CARBOXYL-TERMINAL 7-RESIDUE REPEATS.  
 FT VARIANT 1653 1659 MISSING (IN STRAIN A364A).  
 FT CONFLICT 1514 1514 A -> V (IN REF. 1).  
 FT CONFLICT 1524 1524 G -> A (IN REF. 1).  
 FT CONFLICT 1601 1601 T -> M (IN REF. 1).  
 SQ SEQUENCE 1733 AA; 191610 MW; A45C1360FF99P966 CRC64;

Query Match 100.0%; Score 38; DB 1; Length 1733;  
 Best Local Similarity 100.0%; Pred. No. 25;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YSPSPS 7  
 DB 1556 YSPSPS 1562

RESULT 4  
 RPBL\_SCHPO STANDARD; PRT; 1752 AA.  
 AC P36594;  
 DT 01-JUN-1994 (Rel. 29, Created)  
 DT 01-JUN-1994 (Rel. 29, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE DNA-directed RNA polymerase II largest subunit (EC 2.7.7.6)  
 DE (RNA polymerase II subunit 1).  
 GN RPBI OR SPBC28F2.12.  
 OS Schizosaccharomyces pombe (fission yeast).  
 CC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 CC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 CC Schizosaccharomycetes.  
 CC NCBI\_TaxID=4896;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RX MEDLINE=91187661; PubMed=2011520;  
 RA Azuma Y., Yarnagishi M., Ueshima R., Ishihama A.;  
 RT "Cloning and sequence determination of the Schizosaccharomyces pombe  
 rtp1 gene encoding the largest subunit of RNA polymerase II.",  
 RL Nucleic Acids Res. 19:461-468(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RX MEDLINE=21848401; PubMed=11859360;  
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
 RA Collins K., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,  
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
 RA Woodward J., Volkart G., Aert R., Robben J., Gymnopoulos B.,  
 RA Wellens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Filtz C., Holzer H., Moestl D., Hilbert H.,  
 RA Borzym K., Langer I., Beck A., Leirach E., Reinhardt J., Pohl T.M.,  
 RA Eger P., Zimmermann W., Medler H., Wambutt R., Purnelle B.,  
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaire V., Mortier S.,  
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hust S.M.,  
 RA Lucas M., Rochet M., Gaillardin C., Tallada V.A., Rayon A., Thode G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
 RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
 RA Shpakovski G.V., Ussery D., Barrett B.G., Nurse P.;  
 RT "The genome sequence of Schizosaccharomyces pombe.";  
 RL Nature 415:871-880(2002).

CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION  
 CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS  
 CC SUBSTRATES.  
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
 CC [RNA] (N).  
 CC -1- SUBUNIT: RNA POLYMERASE II CONSISTS OF 10 DIFFERENT SUBUNITS.  
 CC THIS SUBUNIT IS THE LARGEST COMPONENT OF RNA POLYMERASE II.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- PTM: THE TANDUM 7 RESIDUES REPEATS CAN BE HIGHLY PHOSPHORYLATED.  
 CC THE PHOSPHORYLATION ACTIVATES POL2.  
 CC -1- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE  
 CC FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA  
 CC PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE  
 CC III FOR 5S AND 5.8S GENES.  
 CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----

DR EMBL: X5564; CAA39916.1; -;  
 DR EMBL: AL121795; CAB57941.1; -;  
 DR PIR: S26849; S26849.  
 DR InterPro: IPR000684; RNA\_polII\_repeat.  
 DR InterPro: IPR000722; RNA\_pol\_A.  
 DR InterPro: IPR002879; RNA\_pol\_A2.  
 DR InterPro: IPR00623; RNA\_pol\_A1.  
 DR Pfam: PF01854; RNA\_pol\_A2; 1.  
 DR Pfam: PF01854; RNA\_pol\_A1.  
 DR PROSITE: PS00115; RNA\_POL\_II\_REPEAT; 24.  
 KW Transferring: DNA-directed RNA polymerase; Transcription; Zinc; Repeat;  
 KW DNA-binding: Nuclear protein; Phosphorylation; Zinc-finger.  
 FT ZN\_FING 69 85 C2H2-TYPE (POTENTIAL).  
 FT DOMAIN 1554 1752 CARBOXYL-TERMINAL 7-RESIDUE REPEATS.  
 SQ SEQUENCE 1752 AA; 194161 MW; 15A4F0B59E0E570 CRC64;

Query Match 100.0%; Score 38; DB 1; Length 1752;  
 Best Local Similarity 100.0%; Pred. No. 26;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YSPSPS 7  
 DB 1578 YSPSPS 1584

RESULT 5  
 RPBL\_ARATH STANDARD; PRT; 1840 AA.  
 ID RPBL\_ARATH  
 AC P18616; P31635; Q9S258;  
 DT 01-NOV-1990 (Rel. 16, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DE DNA-directed RNA polymerase II largest subunit (EC 2.7.7.6).  
 DE 16-OCT-2001 (Rel. 40, Last annotation update)  
 GN RPB205 OR RPB1 OR RPB1 OR A74G35800 OR PAB14.70.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;  
 RX MEDLINE=91080867; PubMed=2259344;  
 RA Nawrath C., Scheil J., Koncz C.;  
 RT "Homologous domains of the largest subunit of eucaryotic RNA  
 RT polymerase II are conserved in plants.";  
 RL Mol. Gen. Genet. 223:65-75(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Columbia;

RX MEDLINE=91355869; PubMed=2103447;  
 RA Dietrich M.A., Preening J.P., Guilfoyle T.J.;  
 RT "Analysis of the genes encoding the largest subunit of RNA polymerase  
 RT II in Arabidopsis and soybean".  
 RL Plant Mol. Biol. 15:207-223(1990).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV, Columbia;  
 RX MEDLINE=20083488; PubMed=10617198;  
 RA Mayer K.F.X., Scheller C., Manbut R., Murphy G., Volckaert G.,  
 RA Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terryn N.,  
 RA Harris B., Ansoer W., Brandt P., Grivell L.A., Rieger M.,  
 RA Weichselgartner M., de Simone V., Oberwallner B., Mache R., Mueller M.,  
 RA Kreis M., Deiseny M., Puigdomenech P., Watson M., Schmidheini T.,  
 RA Reichert B., Portelle D., Perez-Alonso M., Boutry M., Bancroft I.,  
 RA Vos P., Hohseil J., Zimmermann W., Medler H., Ridley P.,  
 RA Langham S.-A., McCullagh B., Bilham L., Robben J.,  
 RA van der Schueren J., Grynoprez B., Chuang Y.-D., Vandenbussche F.,  
 RA Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,  
 RA Weltenegeger T., Botne G., Ransperger U., Hilbert H., Braun M.,  
 RA Holzer E., Brandt A., Peters S., van Staveren M., Dirkse W.,  
 RA Moolman P., Klein lankhorst R., Rose M., Hauf J., Koeltter P.,  
 RA Bernier S., Hempel S., Feldausch M., Lamberth S., van den Daele H.,  
 RA de Keyser A., Buysaert C., Gielen J., Villarroel R., De Clercq R.,  
 RA van Montagu M., Kogers J., Cronin A., Quail M., Bray-Allen S.,  
 RA Clark L., Doggett J., Hall S., Kay M., Lennard N., McIlroy K., Mayes R.,  
 RA Pettit A., Rajandream M.A., Lyne M., Benes V., Reckmann S.,  
 RA Borova D., Bloeker H., Scharfe M., Grimm M., Loehner T.-H.,  
 RA Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fartmann B., Granderath K., Danner D., Herzl A.,  
 RA Neumann S., Argitlou A., Vitale D., Liguori R., Piravandi E.,  
 RA Massenet O., Gulgley F., Clabaud G., Muendlein A., Felber R.,  
 RA Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,  
 RA Cheffor F., Cooke R., Berger C., Montfort A., Casacuberta E.,  
 RA Gibbons T., Weber N., Vandenbol M., Barges M., Terol J., Torres A.,  
 RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,  
 RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bieleke C.,  
 RA Frisman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,  
 RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,  
 RA Parnell L., Dedha N., Gnoj L., Schutz K., Huang E.,  
 RA Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,  
 RA Stoeckling T., Kalicki J., Graves T., Harmon G., Edwards J.,  
 RA Latelle P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,  
 RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,  
 RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,  
 RA Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,  
 RA Du H., Ali J., Berghoff A., Jones C., Drone K., Cotton M., Joshi C.,  
 RA Antonou B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,  
 RA Ma P., Zhong J., Preston R., Vill D., Shekher M., Matero A., Shan R.,  
 RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,  
 RA Granat S., Shohdy N., Hasegawa A., Hameed A., Lodi M., Johnson A.,  
 RA Chen E., Maria M., Martienssen R., McCombie W.R.;  
 RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis  
 RT thaliana".  
 RT Nature 402:769-777(1999).  
 RL  
 CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION  
 CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS  
 CC SUBSTRATES.  
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
 CC {RNA}(N).  
 CC -1- SUBCELLULAR LOCATION: Nucleus.  
 CC -1- PTM: THE TANDEM 7 RESIDUES REPEATS CAN BE HIGHLY PHOSPHORYLATED.  
 CC THE PHOSPHORYLATION ACTIVATES POL2.  
 CC -1- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES WERE  
 CC FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA  
 CC PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE  
 CC III FOR 5S AND 5.8S GENES.  
 CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.  
 CC  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.lsb-sib.ch/announce/>  
 CC or send an email to [license@lsb-sib.ch](mailto:license@lsb-sib.ch)).  
 CC -----  
 CC DR EMBL; X52954; CAAB37130.1; -  
 CC DR EMBL; X52494; CAAB6735.1; -  
 CC DR EMBL; AL031986; CAA21466.2; -  
 CC DR EMBL; AL015188; CAB01489.1; -  
 CC DR PIR; S12071; JDM01.  
 CC DR PIR; S11960; JDM02.  
 CC DR InterPro; IPR000684; RNA\_PolII\_repeat.  
 CC DR InterPro; IPR000722; RNA\_Pol\_A.  
 CC DR InterPro; IPR002879; RNA\_Pol\_A2.  
 CC DR Pfam; PF00623; RNA\_Pol\_A; 1.  
 CC DR Pfam; PF01854; RNA\_Pol\_A2; 1.  
 CC DR PROSITE; PS00115; RNA\_Pol\_II\_REPEAT; 23.  
 CC KW Transferrase; DNA-directed RNA polymerase; Transcription; Zinc; Repeat;  
 CC RNA-binding; Nucleic acid; Phosphorylation; Zinc-finger.  
 CC FT ZN-FING 66  
 CC FT DNA\_BIND 327 398  
 CC FT DOMAIN 786 796  
 CC FT 1530 1819  
 CC FT 117 124  
 CC FT 193 197  
 CC FT 298 298  
 CC FT 303 303  
 CC FT 401 417  
 CC FT 428 428  
 CC FT 446 446  
 CC FT 739 739  
 CC FT 1062 1062  
 CC FT 1089 1089  
 CC FT 1720 1720  
 CC SO SEQUENCE 1840 AA; 20468 MW; 8453621AD945C1B6 CRC64;  
 CC  
 CC Query Match 100.0%; Score 38; DB 1; Length 1840;  
 CC Best Local Similarity 100.0%; Pred. No. 27;  
 CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 CC  
 CC QY 1 YSPSPS 7  
 CC Db 1594 YSPSPS 1600  
 CC  
 CC RESULT 6  
 CC ID REP1\_CAEEL STANDARD; PRT; 1852 AA.  
 CC AC P16356; Q20090;  
 CC DT 01-AUG-1990 (Rel. 15, Created)  
 CC DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 CC DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 CC DE DNA-directed RNA polymerase II largest subunit (EC 2.7.7.6).  
 CC GN ARA-1 OR F36A4.7.  
 CC OS Caenorhabditis elegans.  
 CC OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 CC OC Rhabditidae; Pelodermidae; Caenorhabditis.  
 CC OX NCBI\_TaxID=6239;  
 CC RN  
 CC RP SEQUENCE FROM N.A.  
 CC RC STRAIN=Bristol N2;  
 CC RA Miller N., Bradshaw H.;  
 CC RL Submitted (Apr-1996) to the EMBL/GenBank/DBJ databases.  
 CC [3]

REVIEWS.

RA Watson R.;

RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS SUBSTRATES.

CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate + (RNA)(N).

CC -1- SUBUNIT: EACH CLASS OF RNA POLYMERASE IS ASSEMBLED FROM NINE TO FOURTEEN DIFFERENT POLYPEPTIDES.

CC -1- SUBCELLULAR LOCATION: Nuclear.

CC -1- PPM: THE TANDEM 7 RESIDUES REPEATS CAN BE HIGHLY PHOSPHORYLATED. THE PHOSPHORYLATION ACTIVATES POL2.

CC -1- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES WERE FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE III FOR 5S AND TRNA GENES.

CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL: M29235; AAA28126.1; -;

DR EMBL: U53333; AAA96158.2; -;

DR PIR: A34092; A34092.

DR Wormpep: F36A4.7; CE28300.

DR Interpro: IPR000684; RNA\_polII\_repeat.

DR Interpro: IPR002879; RNA\_pol\_A2.

DR Interpro: IPR002879; RNA\_pol\_A2.

DR Pfam: PF00623; RNA\_pol\_A; 1.

DR Pfam: PF01854; RNA\_pol\_A2; 1.

DR PROSITE: PS00115; RNA\_POL\_II\_REPEAT; 26.

DR TRANSFAC: DNA-directed RNA polymerase; Transcription; zinc; Repeat; DNA-binding; Nuclear protein; Phosphorylation; zinc-finger.

FT ZN\_FING 66 82 CARBOXYL-TERMINAL 7-RESIDUE REPEATS.

FT DOMAIN 1360 1852

FT CONFLICT 215 215 V -> D (IN REF. 1).

FT CONFLICT 911 911 R -> RVSVAQNAIKL (IN REF. 1).

FT CONFLICT 959 959 I -> D (IN REF. 1).

FT CONFLICT 974 974 O -> L (IN REF. 1).

FT CONFLICT 990 991 KP -> NA (IN REF. 1).

FT CONFLICT 1156 1158 MISSING (IN REF. 1).

FT CONFLICT 1402 1403 IT -> IS (IN REF. 1).

FT SEQUENCE 1852 AA; 203978 MW; 211E4E563119088B CRC64;

Query Match 100.0%; Score 38; DB 1; Length 1852;

Best Local Similarity 100.0%; Pred. No. 27;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YSPSPS 7

DB 1589 YSPSPS 1595

RESULT 7

RBI1\_DROME STANDARD; PRT; 1887 AA.

ID RBI1\_DROME

AC P04052; G9YVX6;

DT 01-NOV-1986 (Rel. 03, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE DNA-directed RNA polymerase II largest subunit (EC 2.7.7.6).

GN RPII215 OR CG1554.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;

OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

OX NCBI\_TaxID=7227;

[1]

RN SEQUENCE FROM N.A.

RP MEDLINE=89218930; PubMed=2496296;

RX Jokest R.S., Weeks J.R., Zehring W.A., Greenleaf A.L.;

RA "Analysis of the gene encoding the largest subunit of RNA polymerase II in *Drosophila*."

RT Mol. Gen. Genet. 215:266-275(1989).

RL [2]

RP SEQUENCE FROM N.A.

RC STRAIN=Berkeley;

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Morten J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champs C., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,

RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,

RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeigwe C.,

RA Jaitani M., Kalush F., Kapen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Jallat B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Styksas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

RA Ye J., Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

RT "The genome sequence of *Drosophila melanogaster*."

RL Science 287:2185-2195(2000).

RN [3]

RP SEQUENCE OF 1-27 FROM N.A.

RX MEDLINE=87089662; PubMed=3025586;

RA Searles L.L., Greenleaf A.L., Kemp M.E., Voelker R.A.;

RT "Sites of P element insertion and structures of P element deletions in the 5' region of *Drosophila melanogaster* RPII215."

RL Mol. Cell. Biol. 6:3312-3319(1986).

RN [4]

RP SEQUENCE OF 1-472 FROM N.A.

RX MEDLINE=85282618; PubMed=2992806;

RA Biggs J., Searles L.L., Greenleaf A.L.;

RT "Structure of the eukaryotic transcription apparatus: features of the gene for the largest subunit of *Drosophila* RNA polymerase II."

RL Cell 42:611-621(1985).

RN [5]

RP SEQUENCE OF 1441-1887 FROM N.A.

RX MEDLINE=88094402; PubMed=3122024;

RA Allison L.A., Wong J.K.-C., Fitzpatrick V.D., Moyle M., Ingles C.J.;

RT "The C-terminal domain of the largest subunit of RNA polymerase II of *Saccharomyces cerevisiae*, *Drosophila melanogaster*, and mammals: a conserved structure with an essential function."

RL Mol. Cell. Biol. 8:321-329(1988).

CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS

CC SUBSTRATES.  
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
 CC (RNA)(N).  
 CC -1- SUBUNIT: EACH CLASS OF RNA POLYMERASE IS ASSEMBLED FROM TEN TO  
 CC TWELVE DIFFERENT POLYPEPTIDES. THE 215 KDA POLYPEPTIDE IS THE  
 CC LARGEST COMPONENT OF RNA POLYMERASE II.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- PTM: THE TANDEM 7 RESIDUES REPEATS CAN BE HIGHLY PHOSPHORYLATED.  
 CC -1- THE PHOSPHORYLATION ACTIVATES POL2.  
 CC -1- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES WERE  
 CC FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA  
 CC PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE  
 CC III FOR 5S AND TRNA GENES.  
 CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: M27431; AAA28868.1; -  
 CC EMBL: AE003486; AAF48057.1; -  
 CC EMBL: M14203; AAA28864.1; -  
 CC EMBL: M11798; AAA28863.1; -  
 CC EMBL: M19537; AAA28827.1; -  
 CC PIR: S04457; RNF2L  
 CC FlyBase: FBgn0003277; Rp11215.  
 CC InterPro: IPR000684; RNA\_polII\_repeat.  
 CC InterPro: IPR000722; RNA\_pol\_A.  
 CC InterPro: IPR002879; RNA\_pol\_A2.  
 CC Pfam: PF00623; RNA\_pol\_A; 1.  
 CC Pfam: PF01854; RNA\_pol\_A2; 1.  
 CC PROSITE: PS00115; RNA\_POL\_II\_REPEAT; 11.  
 CC Transfaser: DNA-directed RNA polymerase; Transcription; Zinc; Repeat;  
 CC DNA-binding; Nuclear protein; Phosphorylation; Zinc-finger.  
 CC ZN\_FING 67 83  
 CC DOMAIN 1579 1881 C2H2-TYPE (POTENTIAL).  
 CC CONFLICT 319 324 CARBOXYL-TERMINAL 7-RESIDUE REPEATS.  
 CC CONFLICT 450 450 P -> G (IN REF. 4).  
 CC CONFLICT 455 458 TLHK -> RCTT (IN REF. 4).  
 CC CONFLICT 463 472 GHRVRLPWS -> VTGSVASST (IN REF. 4).  
 CC CONFLICT 741 741 R -> H (IN REF. 1).  
 CC CONFLICT 1485 1524 SMLGGAFFIGSGSPSWTPTWANCNTPRYSPGHV  
 CC -> I (IN REF. 5).  
 CC CONFLICT 1506 1508 MTP -> ELDSA (IN REF. 1).  
 CC CONFLICT 1887 1887 D -> DVKGGRG (IN REF. 1).  
 CC SEQUENCE 1887 AA; 209167 MW; 4EC68C7708A167A3 CRC64;  
 SQ  
 Query Match 100.0%; Score 38; DB 1; Length 1887;  
 Best Local Similarity 100.0%; Pred. No. 28;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 YSPTSPS 7  
 DB 1685 YSPTSPS 1691  
 RESULT 8  
 RPB1\_HUMAN STANDARD; PRT; 1970 AA.  
 AC P24928;  
 DT 01-MAR-1992 (Rel. 21, Created)  
 DT 01-MAR-1992 (Rel. 21, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE DNA-directed RNA polymerase II largest subunit (EC 2.7.7.6) (RPB1).  
 GN POLR2A.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OC NCBI\_TaxId=9606;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-92178992; PubMed-1542581.  
 RA Wintzerith M., Acker J., Vicaire S., Vigneron M., Keding C.;  
 RT "Complete sequence of the human RNA polymerase II largest subunit.";  
 RL Nucleic Acids Res. 20:910-910(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-95347616; PubMed-7622068;  
 RA Mita K., Tsuji H., Morimyo M., Takahashi E., Nenoi M.,  
 RA Ichimura S., Yamauchi M., Hongo E., Hayashi A.;  
 RT "The human gene encoding the largest subunit of RNA polymerase II.";  
 RL Gene 159:285-286(1995).  
 CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION  
 CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS  
 CC SUBSTRATES.  
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
 CC (RNA)(N).  
 CC -1- SUBUNIT: RNA POLYMERASE II CONSISTS OF 12 DIFFERENT SUBUNITS.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- PTM: THE TANDEM 7 RESIDUES REPEATS CAN BE HIGHLY PHOSPHORYLATED.  
 CC -1- THE PHOSPHORYLATION ACTIVATES POL2.  
 CC -1- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE  
 CC FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA  
 CC PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE  
 CC III FOR 5S AND TRNA GENES.  
 CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X63564; CAA45125.1; -  
 CC EMBL: X74874; CAA52862.1; JOINED.  
 CC EMBL: X74873; CAA52862.1; JOINED.  
 CC EMBL: X74872; CAA52862.1; JOINED.  
 CC EMBL: X74871; CAA52862.1; JOINED.  
 CC EMBL: X74870; CAA52862.1; JOINED.  
 CC PIR: S21054; S21054.  
 CC GeneW: HGNC:9187; POLR2A.  
 CC MIM: 180660; -  
 CC InterPro: IPR000684; RNA\_polII\_repeat.  
 CC InterPro: IPR000722; RNA\_pol\_A.  
 CC InterPro: IPR002879; RNA\_pol\_A2.  
 CC Pfam: PF00623; RNA\_pol\_A; 1.  
 CC Pfam: PF01854; RNA\_pol\_A2; 1.  
 CC PROSITE: PS00115; RNA\_POL\_II\_REPEAT; 43.  
 CC Transfaser: DNA-directed RNA polymerase; Transcription; Zinc; Repeat;  
 CC DNA-binding; Nuclear protein; Phosphorylation; Zinc-finger.  
 CC ZN\_FING 71 87  
 CC DOMAIN 1590 1958 CARBOXYL-TERMINAL 7-RESIDUE REPEATS.  
 CC CONFLICT 1067 1067 W -> L (IN REF. 2).  
 CC CONFLICT 1449 1449 D -> Y (IN REF. 2).  
 CC SEQUENCE 1970 AA; 217205 MW; 6876FC25692A657E CRC64;  
 SQ  
 Query Match 100.0%; Score 38; DB 1; Length 1970;  
 Best Local Similarity 100.0%; Pred. No. 29;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 YSPTSPS 7  
 DB 1615 YSPTSPS 1621  
 RESULT 9  
 RPB1\_MOUSE STANDARD; PRT; 1970 AA.  
 AC P08775;  
 DT 01-NOV-1988 (Rel. 09, Created)



01-MAR-1992 (Rel. 21, last sequence update)  
 16-OCT-2001 (Rel. 40, last annotation update)  
 DNA-directed RNA polymerase II largest subunit (EC 2.7.7.6) (RPB1).  
 GN POLR2A OR RP02-1 OR RPII215.  
 OS Mus musculus (Mouse).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 CC NCBI\_TaxID=10090;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RX MEDLINE=87280135; PubMed=3038894;  
 RA Ahern J.M., Jr., Bartolomei M.S., West M.L., Cisek L.J., Corden J.L.;  
 RT "Cloning and sequence analysis of the mouse genomic locus encoding  
 RT the largest subunit of RNA polymerase II.";  
 RL J. Biol. Chem. 262:10695-10705(1987).  
 RN [2]  
 RN SEQUENCE OF 1587-1970 FROM N.A.  
 RX MEDLINE=86068017; PubMed=2999785;  
 RA Corden J.L., Cadena D.L., Ahern J.M., Jr., Dahmus M.E.;  
 RT "A unique structure at the carboxyl terminus of the largest subunit  
 RT of eukaryotic RNA polymerase II.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 82:7934-7938(1985).  
 RN [3]  
 RN REVISIONS, AND PRESENCE OF AN ADDITIONAL EXON.  
 RX MEDLINE=92178992; PubMed=1542581;  
 RA Wintzerith M., Acker J., Vicaire S., Vigneron M., Keding C.;  
 RT "Complete sequence of the human RNA polymerase II largest subunit.";  
 CC Nucleic Acids Res. 20:910-910(1992).  
 CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION  
 CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS  
 CC SUBSTRATES.  
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
 CC (RNA)(N).  
 CC -1- SUBUNIT: RNA POLYMERASE II CONSISTS OF 12 DIFFERENT SUBUNITS.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- PTM: THE TANDEM 7 RESIDUES REPEATS CAN BE HIGHLY PHOSPHORYLATED.  
 CC THE PHOSPHORYLATION ACTIVATES POL2.  
 CC -1- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE  
 CC FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA  
 CC PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE  
 CC III FOR 5S AND TRNA GENES.  
 CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M12130; AAA40071.1; -  
 DR EMBL: M14101; AAA40071.1; JOINED.  
 DR PIR: A28490; A28490.  
 DR MGD: MGI:98086; RPO2-1.  
 DR InterPro: IPR000684; RNA\_poliI\_repeat.  
 DR InterPro: IPR000722; RNA\_poliI\_A2.  
 DR InterPro: IPR002879; RNA\_poliI\_A2.  
 DR Pfam: PF00623; RNA\_pol\_A; 1.  
 DR Pfam: PF01854; RNA\_pol\_A2; 1.  
 DR PROSITE: PS00115; RNA\_POL\_II\_REPEAT; 42.  
 KW Transferrase: DNA-directed RNA polymerase; Transcription; Zinc; Repeat;  
 KW DNA-binding; Nuclear protein; Phosphorylation; Zinc-finger.  
 FT ZN\_FING 71 87  
 FT DOMAIN 1590 1958  
 FT CONFLICT 1498 1498 P -> R (IN REF. 1 AND 2).  
 FT CONFLICT 1499 1536 MISSING (IN REF. 1 AND 2).  
 FT SEQUENCE 1970 AA; 217175 MW; 7D6F38FD92A657E CRC64;  
 Query Match 100.0%; Score 38; DB 1; Length 1970;  
 Best Local Similarity 100.0%; Pired. No. 29;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YSPTSPS 7  
 ID 1615 YSPTSPS 1621  
 Db 1615 YSPTSPS 1621  
 RESULT 10  
 RPB1\_PLAFD STANDARD; PRT; 2452 AA.  
 AC P14248;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DE 15-DEC-1998 (Rel. 37, Last annotation update)  
 DT DNA-directed RNA polymerase II largest subunit (EC 2.7.7.6).  
 GN RPII  
 OS Plasmodium falciparum (isolate CDC / Honduras).  
 CC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.  
 CC NCBI\_TaxID=5836;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RX MEDLINE=90098832; PubMed=2690004;  
 RA Li W.B., Beik D.J., Gu H., Tanaka M., Fox B.A., Inselburg J.;  
 RT "An enlarged largest subunit of Plasmodium falciparum RNA polymerase  
 RT II defines conserved and variable RNA polymerase domains.";  
 RL Nucleic Acids Res. 17:9621-9636(1989).  
 CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION  
 CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS  
 CC SUBSTRATES.  
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +  
 CC (RNA)(N).  
 CC -1- SUBUNIT: EACH CLASS OF RNA POLYMERASE IS ASSEMBLED FROM TEN TO  
 CC TWELVE DIFFERENT POLYPEPTIDES. THIS POLYPEPTIDE IS THE LARGEST  
 CC COMPONENT OF RNA POLYMERASE II.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- PTM: THE TANDEM 7 RESIDUES REPEATS CAN BE HIGHLY PHOSPHORYLATED.  
 CC THE PHOSPHORYLATION ACTIVATES POL2.  
 CC -1- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES WERE  
 CC FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA  
 CC PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE  
 CC III FOR 5S AND TRNA GENES.  
 CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X16561; CAA34560.1; -  
 DR PIR: S07485; RNZ02L.  
 DR InterPro: IPR000684; RNA\_poliI\_repeat.  
 DR InterPro: IPR000722; RNA\_pol\_A.  
 DR InterPro: IPR002879; RNA\_pol\_A2.  
 DR Pfam: PF00623; RNA\_pol\_A; 1.  
 DR Pfam: PF01854; RNA\_pol\_A2; 1.  
 DR PROSITE: PS00115; RNA\_POL\_II\_REPEAT; 9.  
 KW Transferrase: DNA-directed RNA polymerase; Transcription; Zinc; Repeat;  
 KW DNA-binding; Nuclear protein; Phosphorylation; Zinc-finger.  
 FT ZN\_FING 68 84  
 FT DOMAIN 1590 1958  
 FT CONFLICT 1498 1498 P -> R (IN REF. 1 AND 2).  
 FT CONFLICT 1499 1536 MISSING (IN REF. 1 AND 2).  
 FT SEQUENCE 1970 AA; 217175 MW; 7D6F38FD92A657E CRC64;  
 Query Match 100.0%; Score 38; DB 1; Length 1970;  
 Best Local Similarity 100.0%; Pired. No. 29;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

SQ SEQUENCE 2452 AA; 278166 MW; F995E117F617A48F CRC64;  
 Query Match 92.1%; Score 35; DB 1; Length 2452;  
 Best Local Similarity 85.7%; Pred. No. 1.3e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 YSPSPS 7  
 |||||  
 Db 2251 YSPSPSPT 2257

RESULT 11  
 YLBT\_CAEEL STANDARD; PRT; 192 AA.  
 AC P46581;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Hypothetical 21.9 kDa protein C34E10.7 in chromosome III.  
 GN C34E10.7.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Pelodierinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RP STRAIN-Bristol N2;  
 RA Kirsten J.;  
 RL Submitted (JUN-1994) to the EMBL/Genbank/DDSI databases.  
 CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).  
 CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF  
 CC TRANSCRIPTION FACTORS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U010402; AAA19069.1; -  
 DR TRANSFAC: T01624; -  
 DR WormPep: C34E10.7; CE01187.  
 DR InterPro: IPR001092; HLH\_basic.  
 DR Pfam: PF00010; HLH\_1.  
 DR SMART: SM00353; HLH\_1.  
 DR SMART: PS00038; HLH\_1; 1.  
 DR PROSITE: PS00888; HLH\_2; 1.  
 DR PROSITE: PS00888; HLH\_2; 1.  
 KW Hypothetical protein; Nuclear protein; DNA-binding.  
 FT DNA\_BIND 22 31 BASIC DOMAIN.  
 FT DOMAIN 32 72 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).  
 FT DOMAIN 139 151 POLY-SER.  
 SQ SEQUENCE 192 AA; 21845 MW; B1BCFEAF6C7310F1 CRC64;

Query Match 89.5%; Score 34; DB 1; Length 192;  
 Best Local Similarity 85.7%; Pred. No. 13;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 YSPSPS 7  
 |||||  
 Db 157 YSPSPS 163

RESULT 12  
 VIE2\_NPVOP STANDARD; PRT; 405 AA.  
 AC P32511;  
 DT 01-OCT-1993 (Rel. 27, Created)  
 DT 01-OCT-1993 (Rel. 27, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Immediate-early protein IE-2 (Trans-activator IE-2).  
 GN IE-2.

OS Orgyia pseudotsugata multicausid polyhedrosis virus (OpMV).  
 CC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;  
 CC Nucleopolyhedrovirus.  
 OX NCBI\_TaxID=164623;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92142536; PubMed=1736546;  
 RA Theilmann D.A.; Stewart S.;  
 RT "Molecular analysis of the trans-activating IE-2 gene of Orgyia  
 RT pseudotsugata multicausid nuclear polyhedrosis virus.";  
 RL Virology 187:84-96(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97271300; PubMed=9126251;  
 RA Ahrens C.H.; Russell R.R.; Funk C.J.; Evans J.; Harwood S.;  
 RA Rohrmann G.F.;  
 RT "The sequence of the Orgyia pseudotsugata multinucleocapsid nuclear  
 RT polyhedrosis virus genome.";  
 RL Virology 229:381-399(1997).  
 CC -1- FUNCTION: TRANSACTIVATES THE IE-1 GENE.  
 CC -1- SIMILARITY: TO ACMPV IMMEDIATE-EARLY REGULATORY PROTEIN IE-N.  
 CC -1- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M83827; AAA6749.1; -  
 DR EMBL: U75930; AAC59150.1; -  
 DR PIR: A42190; WMNVPL.  
 DR InterPro: IPR001841; Znf\_fing.  
 DR Pfam: PF00097; Zf-C3HC4; 1.  
 DR SMART: SM00184; RING\_1.  
 DR PROSITE: PS00518; ZF\_RING\_1; 1.  
 DR PROSITE: PS00518; ZF\_RING\_2; 1.  
 KW Early protein; Transcription regulation; Trans-acting factor;  
 KW Activator; Zinc-finger; DNA-binding.  
 FT DOMAIN 3 53 ARG-RICH (BASIC).  
 FT ZN\_FING 207 255 RING-TYPE.  
 SQ SEQUENCE 405 AA; 45669 MW; 4272E7B2875F85CD CRC64;

Query Match 89.5%; Score 34; DB 1; Length 405;  
 Best Local Similarity 100.0%; Pred. No. 25;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YSPSPS 6  
 |||||  
 Db 140 YSPSPS 145

RESULT 13  
 ZFH2\_DROME STANDARD; PRT; 3005 AA.  
 ID ZFH2\_DROME  
 AC P28167;  
 DT 01-OCT-1994 (Rel. 30, Created)  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 01-OCT-1994 (Rel. 30, Last annotation update)  
 DE Zinc-finger protein 2 (Zinc-finger homeodomain protein 2).  
 GN Zfh-2.  
 OS Drosophila melanogaster (Fruit fly).  
 CC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;  
 CC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 CC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92001539; PubMed=1680376;  
 RA Fortini M.E.; Lai Z.; Rubin G.M.;  
 RA "The Drosophila zfh-1 and zfh-2 genes encode novel proteins

RT containing both zinc-finger and homeodomain motifs.";  
 RL Mech. Dev. 34:113-122(1991).  
 CC -I- FUNCTION: INVOLVED IN THE DEVELOPMENT OF THE EMBRYONIC CENTRAL  
 CC NERVOUS SYSTEM.  
 CC -I- SUBCELLULAR LOCATION: Nuclear (Probable).  
 CC -I- TISSUE SPECIFICITY: LARGELY RESTRICTED TO THE CNS OF LATE EMBRYO.  
 CC -I- SIMILARITY: CONTAINS 3 HOMEODOMAIN DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M63450: AAA29051.1; -.  
 DR PIR: S27817; S27817.  
 DR PIR: S33642; S33642.  
 DR HSSP: P15822; 32NF.  
 DR TRANSFAC: T00920; -.  
 DR FLYBASE: FBgn0004607; zfh2.  
 DR InterPro: IPR001356; Homeobox.  
 DR InterPro: IPR000822; Znf\_C2H2.  
 DR InterPro: IPR000690; Znf\_Matrin.  
 DR Pfam: PF00046; homeobox; 3.  
 DR Pfam: PF00096; zf\_C2H2; 12.  
 DR ProDom: PD000010; Homeobox; 3.  
 DR SMART: SM00389; HOX; 3.  
 DR SMART: SM00355; Znf\_C2H2; 14.  
 DR PROSITE: PS00027; HOMEODOM\_1; 2.  
 DR PROSITE: PS00071; HOMEODOM\_2; 3.  
 DR PROSITE: PS00028; ZINC\_FINGER\_C2H2\_1; 8.  
 DR PROSITE: PS00157; ZINC\_FINGER\_C2H2\_2; 7.  
 KW Zinc-finger; Metal-binding; DNA-binding; Homeobox; Nuclear protein;  
 KW Repeat.  
 FT ZN\_FING 133 156 C2H2-TYPE.  
 FT ZN\_FING 559 582 C2H2-TYPE.  
 FT ZN\_FING 614 638 C2H2-TYPE.  
 FT ZN\_FING 732 756 C2H2-TYPE.  
 FT ZN\_FING 897 916 C2H2-TYPE (DEGENERATE).  
 FT ZN\_FING 940 964 C2H2-TYPE.  
 FT ZN\_FING 999 1023 C2H2-TYPE.  
 FT ZN\_FING 1074 1098 C2H2-TYPE.  
 FT ZN\_FING 1210 1233 C2H2-TYPE.  
 FT ZN\_FING 1341 1365 C2H2-TYPE.  
 FT ZN\_FING 1438 1462 C2H2-TYPE.  
 FT ZN\_FING 1477 1500 C2H2-TYPE (DEGENERATE).  
 FT ZN\_FING 1513 1535 C2H2-TYPE.  
 FT ZN\_FING 1541 1564 C2H2-TYPE.  
 FT ZN\_FING 1797 1856 HOMEODOM\_1.  
 FT DNA\_BIND 2154 2213 HOMEODOM\_2.  
 FT ZN\_FING 2234 2256 C2H2-TYPE.  
 FT ZN\_FING 2371 2393 C2H2-TYPE.  
 FT DNA\_BIND 2760 2819 HOMEODOM\_3.  
 SQ SEQUENCE 3005 AA; 332056 MW; 7E402052BBC6847 CRC64;  
 Query Match 89.5%; Score 34; DB 1; Length 3005;  
 Best Local Similarity 85.7%; Pred. No. 2.5e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 YSPSPSPS 7  
 Db 1174 FSPSPSPS 1180

RESULT 14  
 COAT\_TCV STANDARD; PRT; 351 AA.  
 ID COAT\_TCV  
 AC P06663;  
 DT 01-JAN-1988 (Rel. 06, Created)  
 DT 01-JAN-1988 (Rel. 06, Last sequence update)  
 DT 01-NOV-1995 (Rel. 32, Last annotation update)

DE Coat protein.  
 OS Turnip crinkle virus (TCV).  
 CC Viruses; ssRNA positive-strand viruses, no DNA stage; Tombusviridae;  
 CC Carmovirus.  
 CC NCBI\_TaxID=11988;  
 RN  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=67283926; PubMed=3612806;  
 RA Stockley P.G., Morris T.J.;  
 RT "Structure and assembly of turnip crinkle virus. IV. Analysis of the  
 RT coat protein gene and implications of the subunit primary  
 RT structure.";  
 RL J. Mol. Biol. 194:265-276(1987).  
 RN  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89243179; PubMed=2718381;  
 RA Carrington J.C., Heaton E.A., Zuidema D., Hillman B.I., Morris T.J.;  
 RT "The genome structure of turnip crinkle virus.";  
 RL Virology 170:219-226(1989).  
 RN  
 RP X-RAY CRYSTALLOGRAPHY (3.2 ANGSTROMS).  
 RX MEDLINE=87112726; PubMed=3806676;  
 RA Hogle J.M., Maeda A., Harrison S.C.;  
 RT "Structure and assembly of turnip crinkle virus. I. X-ray  
 RT crystallographic structure analysis at 3.2-A resolution.";  
 RL J. Mol. Biol. 191:625-638(1986).  
 CC -I- SIMILARITY: COAT PROTEINS FROM CARMOVIRUSES (CARMV, MNSV, TCV),  
 CC DIAPHYOVIRUSES (RCNMV), SOBMOVIRUSES (MCMV, SMNV), TOMBSVIRUSES  
 CC (AMCV, CNV, CYRSV, AND TBSV), AND TNV ARE RELATED IN THEIR S  
 CC REGION.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X05193; CAA28823.1; -.  
 DR EMBL: M22445; AAA96971.1; -.  
 DR PIR: JA0111; VCVERTC.  
 DR PIR: S07285; S07285.  
 DR HSSP: P11795; 2TBV.  
 DR InterPro: IPR000937; Viral\_coat.  
 DR Pfam: PF00729; Viral\_coat; 1.  
 DR PRINTS: PR00233; ICOSAHEDRAL.  
 DR PROSITE: PS00355; ICOSAH\_VIR\_COAT\_5; 1.  
 KW Coat protein.  
 FT DOMAIN 1 80 R DOMAIN, INTERACTION WITH RNA.  
 FT DOMAIN 82 238 S DOMAIN, VIRION SHELL.  
 FT DOMAIN 239 351 P DOMAIN, PROJECTING.  
 FT CONFLICT 252 252 D -> Q (TN REF. 2)  
 SQ SEQUENCE 351 AA; 38052 MW; 027C9C2020ECADDC CRC64;  
 Query Match 84.2%; Score 32; DB 1; Length 351;  
 Best Local Similarity 85.7%; Pred. No. 59;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 YSPSPSPS 7  
 Db 138 YSPSPSPS 144

RESULT 15  
 FXHL\_MOUSE STANDARD; PRT; 401 AA.  
 ID FXHL\_MOUSE  
 AC O88621; Q9QZL5; Q9R241;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Forkhead box protein H1 (Forkhead activin signal transducer 1) (Fast-  
 DE 1) (Forkhead activin signal transducer 2) (Fast-2).

GN FOXH1 OR FAST1 OR FAST2.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A. (LONG ISOFORM).  
RA MEDLINE=98367553; PubMed=9702197.  
RX Labbe E., Silvestri C., Hoodless P.A., Wraana J.L., Attisano L.;  
RT "Smad2 and Smad3 positively and negatively regulate TGF beta-dependent  
transcription through the forkhead DNA-binding protein FAST2.";  
RL Mol. Cell 2:109-120(1998).  
RN [2]  
RP SEQUENCE FROM N.A. (LONG ISOFORM).  
RX TISSUE=Embryonic carcinoma;  
RA MEDLINE=99077980; PubMed=9858566;  
RX Liu B., Dou C.-L., Prabhu L., Lai E.;  
RT "FAST-2 is a mammalian winged-helix protein which mediates  
transforming growth factor beta signals.";  
RL Mol. Cell. Biol. 19:424-430(1999).  
RN [3]  
RP SEQUENCE FROM N.A. (LONG AND SHORT ISOFORMS), AND CHARACTERIZATION.  
RX TISSUE=Embryo;  
RA MEDLINE=99279234; PubMed=10349617;  
RX Weisberg E., Winkler G.E., Chen X., Farnsworth C.L., Hogan B.L.H.,  
RA Whitman M.;  
RT "A mouse homologue of FAST-1 transduces TGF beta superfamily signals  
and is expressed during early embryogenesis.";  
RL Mech. Dev. 79:17-27(1998).  
RN [4]  
RP SEQUENCE FROM N.A. (LONG ISOFORM).  
RA Chen Y., Nagarajan R.P., Liu J., Vale M.;  
RT "Mouse FAST-2 transduces the signals for TGF-beta and activin.";  
RL Submitted (DEC-1998) to the EMBL/Genbank/DBJ databases.  
CC -1- FUNCTION: TRANSCRIPTIONAL ACTIVATOR. RECOGNIZES AND BINDS TO THE  
DNA SEQUENCE 5'-TGT[GT][GT]AT-3'. REQUIRED FOR INDUCTION OF THE  
GOOSECOID (GSC) PROMOTER BY TGF-BETA OR ACTIVIN SIGNALING. FORMS A  
TRANSCRIPTIONALLY ACTIVE COMPLEX CONTAINING FAST-1/SMAD2/SMAD4 ON  
A SITE ON THE GSC PROMOTER CALLED TARE (TGF-BETA/ACTIVIN RESPONSE  
ELEMENT).  
CC -1- SUBCELLULAR LOCATION: Nuclear.  
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A  
SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING. THE SHORT  
ISOFORM FAILS TO BIND DNA AND CANNOT CONFER ACTIVIN RESPONSE  
ELEMENT RESPONSIVENESS.  
CC -1- DEVELOPMENTAL STAGE: EXPRESSED PREDOMINANTLY THROUGHOUT THE  
EPIBLAST BEFORE GASTRULATION AND DECLINES AS DEVELOPMENT  
PROGRESSES.  
CC -1- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation  
at the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL; AF069303; AAC79808.1; -;  
DR EMBL; AF079514; AAD12486.1; -;  
DR EMBL; AF177770; AAD55949.1; -;  
DR EMBL; AF110506; AAD14683.1; -;  
DR HSSP; 063245; 2HPH.  
DR TRANSFAC; T04100; -;  
DR TRANSFAC; T04120; -;  
DR MGD; MGI:1347465; Foxh1.  
DR InterPro; IPR001766; TF\_Fork\_head.  
DR Pfam; PF00250; Fork\_head; 1.  
DR PRINTS; PR00053; FORKHEAD.  
DR ProDom; PD000425; TF\_Fork\_head; 1.  
DR SMART; SM00339; FH; 1.  
DR PROSITE; PS00657; FORK\_HEAD\_1; FALSE\_NEG.  
DR PROSITE; PS00658; FORK\_HEAD\_2; FALSE\_NEG.

DR PROSITE; PS00658; FORK\_HEAD\_2; 1.  
KW Transcription regulation; Activator; DNA-binding; Nuclear protein;  
FT DNA\_BIND 64  
FT DOMAIN 307 390  
FT DOMAIN 187 194  
FT DOMAIN 243 246  
FT VARSPPLIC 89 92  
FT CONFLICT 167 167  
FT CONFLICT 207 207  
FT CONFLICT 215 215  
FT CONFLICT 220 220  
FT CONFLICT 223 223  
FT CONFLICT 237 237  
FT CONFLICT 330 330  
FT CONFLICT 334 334  
FT CONFLICT 340 340  
FT CONFLICT 366 366  
SQ SEQUENCE 401 AA; 44001 MW; FB1IDC9779DF1454 CRC64;  
Query Match 84.28; Score 32; DB 1; Length 401;  
Best Local Similarity 71.48;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
Oy 1 YSPSPS 7  
|:|:|:|  
Db 11 YTPPTPS 17  
Search completed: June 23, 2003, 16:17:38  
Job time : 12 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 23, 2003, 16:15:01 ; Search time 29 Seconds  
(without alignments)  
49.736 Million cell updates/sec

Title: US-09-733-773B-3  
Perfect score: 38  
Sequence: 1 YSPSPS 7

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues  
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL\_21:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_protent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	38	100.0	314 10 09F154	09F154 arabidopsis
2	38	100.0	414 10 09A231	09A231 kluyveromy
3	38	100.0	494 10 099367	099367 glycine max
4	38	100.0	625 10 099366	099366 glycine max
5	38	100.0	977 10 099368	099368 glycine max
6	38	100.0	1429 5 08SXM3	08SXM3 drosophila
7	38	100.0	1569 5 000820	000820 acanthocob
8	38	100.0	1599 5 08SSC4	08SSC4 acanthocob
9	38	100.0	1603 3 09C428	09C428 monolephar
10	38	100.0	1605 5 096446	096446 vairimorpha
11	38	100.0	1685 10 004142	004142 porphyra ye
12	38	100.0	1741 10 09FRB0	09FRB0 cryza sativ
13	38	100.0	1743 5 09BMP3	09BMP3 monosiga br
14	38	100.0	1746 10 09AXN0	09AXN0 glaucosphae
15	38	100.0	1748 3 09HGZ5	09HGZ5 aspergillus
16	38	100.0	1811 5 09XZ08	09XZ08 drosophila

17	38	100.0	1889 5 09XZS2	09XZS2 drosophila
18	38	100.0	1889 5 09XZS1	09XZS1 drosophila
19	38	100.0	1889 5 09XZU9	09XZU9 drosophila
20	38	100.0	1966 11 008847	008847 mus musculus
21	38	100.0	1970 11 035559	035559 cricetus
22	35	92.1	204 5 025655	025655 plasmodium
23	35	92.1	455 16 09AAS0	09AAS0 caulobacter
24	35	92.1	1553 5 096452	096452 nosema locu
25	35	92.1	1712 10 09AXN1	09AXN1 botryocodi
26	35	92.1	2457 5 077375	077375 plasmodium
27	34	89.5	345 16 096P25	096P25 rhizobium 1
28	34	89.5	348 2 09Z6D3	09Z6D3 haemophilus
29	34	89.5	517 12 09YU51	09YU51 turkey aden
30	34	89.5	619 5 018266	018266 caenorhabdi
31	34	89.5	628 3 094732	094732 schizosacch
32	34	89.5	1194 3 093962	093962 glomus vers
33	34	89.5	1553 5 08T219	08T219 dictyostell
34	34	89.5	1732 5 009451	009451 bonnemaiso
35	34	89.5	3005 5 09VAD7	09VAD7 drosophila
36	33	86.8	733 3 09P8W3	09P8W3 trichoderma
37	32	84.2	53 16 09PE01	09PE01 xyella fas
38	32	84.2	118 10 08S3A7	08S3A7 physcomitre
39	32	84.2	168 3 094086	094086 saccharomyc
40	32	84.2	353 16 08X307	08X307 escherichia
41	32	84.2	428 10 09SVM9	09SVM9 arabidopsis
42	32	84.2	583 4 060291	060291 homo sapien
43	32	84.2	590 12 0919L7	0919L7 cullex nigri
44	32	84.2	642 2 0937D7	0937D7 oenococcus
45	32	84.2	642 2 009029	009029 oenococcus

## ALIGNMENTS

RESULT 1  
ID 09F154 PRELIMINARY; PRT; 314 AA.  
AC 09F154;  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE Genomic DNA, chromosome 5, TAC clone:K3K7.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons; Rosidae;  
OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=5702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=COLUMBIA;  
RX MEDLINE=99397451; PubMed=10470850;  
RA Kaneko T., Katoh T., Sato S., Nakamura Y., Asamizu E., Kotani H.,  
RA Miyajima N., Tabata S.,  
RT "Structural analysis of Arabidopsis thaliana chromosome 5. IX.  
RT Sequence features of the regions of 1,011,550 bp covered by seventeen  
RT pl and TAC clones.";  
RL DNA Res. 6:183-195(1999).  
DR EMBL: AB017063; BAB08740.1; -;  
DR InterPro: IPR000684; RNA\_POL\_II\_REPEAT. UNKNOWN.1.  
DR PROSITE: PS00115; RNA\_POL\_II\_REPEAT. UNKNOWN.1.  
SQ SOURCE 314 AA: 34579 MW: 3850E794E95BE54D CRC64:

Query Match 100.0%; Score 38; DB 10; Length 314;  
Best Local Similarity 100.0%; Pred No. 11;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YSPSPS 7  
Db 113 YSPSPS 119  
RESULT 2  
094231

ID 094231 PRELIMINARY; PRT: 414 AA.  
AC 094231;  
DT 01-MAY-1999 (TReMBLrel. 10, Created)  
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)  
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)  
DE RNA polymerase II largest subunit (Fragment).  
OS Kluveromyces lactic (Yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Kluveromyces.  
OX NCBI\_TaxID=28985;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Lee J.M., Greenleaf A.L.;  
RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.  
DR EMBL: U24217; AAC99803.1; -  
DR InterPro: IPR002965; P\_rich\_extensn.  
DR InterPro: IPR000684; RNA\_polII\_repeat.  
DR InterPro: IPR002879; RNA\_pol\_A2.  
DR Pfam: PF01854; RNA\_pol\_A2; 1.  
DR PRINTS: PR01217; PRICHEXTENS.  
DR PROSITE: PS00115; RNA\_POL\_II\_REPEAT; UNKNOWN\_20.  
FT NON\_TER 1  
SQ SEQUENCE 414 AA; 43369 MW; 59A11389369B00B3 CRC64;  
  
Query Match 100.0%; Score 38; DB 3; Length 414;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 YSPSPS 7  
DB 244 YSPSPS 250  
  
RESULT 3  
ID 099367 PRELIMINARY; PRT: 494 AA.  
AC 099367;  
DT 01-NOV-1996 (TReMBLrel. 01, Created)  
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)  
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)  
DE DNA-directed RNA polymerase (EC 2.7.7.6) (Fragment).  
GN RPB1-B2.  
OS Glycine max (Soybean).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eucosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.  
OX NCBI\_TaxID=3847;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Dietrich M.A.;  
RL Submitted (MAR-1990) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE OF 204-494 FROM N.A.  
RC STRAIN-VAR.WAYNE;  
RC MEDLINE=91355869; PubMed=2103447;  
RA Dietrich M.A., Prenger J., Guilfoyle T.J.;  
RT Analysis of the genes encoding the largest subunit of RNA polymerase II in Arabidopsis and soybean.";  
RT Plant Mol. Biol. 15:207-223(1990).  
DR EMBL: X52493; CA36734.1; -  
DR InterPro: IPR002965; P\_rich\_extensn.  
DR InterPro: IPR000684; RNA\_polII\_repeat.  
DR InterPro: IPR002879; RNA\_pol\_A2.  
DR Pfam: PF01854; RNA\_pol\_A2; 1.  
DR PRINTS: PR01217; PRICHEXTENS.  
DR PROSITE: PS00115; RNA\_POL\_II\_REPEAT; UNKNOWN\_22.  
KW DNA-directed RNA polymerase; Nucleotidyltransferase; Transferase.  
FT NON\_TER 1  
SQ SEQUENCE 494 AA; 52156 MW; 41EEF0D30F61D649 CRC64;  
  
Query Match 100.0%; Score 38; DB 10; Length 494;  
Best Local Similarity 100.0%; Pred. No. 17;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 YSPSPS 7  
DB 266 YSPSPS 272  
  
RESULT 4  
ID 099366 PRELIMINARY; PRT: 625 AA.  
AC 099366;  
DT 01-NOV-1996 (TReMBLrel. 01, Created)  
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)  
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)  
DE DNA-directed RNA polymerase (EC 2.7.7.6) (Fragment).  
GN RPB1-B1 OR GENE B1.  
OS Glycine max (Soybean).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eucosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.  
OX NCBI\_TaxID=3847;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Dietrich M.A.;  
RL Submitted (MAR-1990) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP PARTIAL SEQUENCE FROM N.A.  
RC STRAIN-VAR.WAYNE;  
RC MEDLINE=91355869; PubMed=2103447;  
RA Dietrich M.A., Prenger J., Guilfoyle T.J.;  
RT Analysis of the genes encoding the largest subunit of RNA polymerase II in Arabidopsis and soybean.";  
RT Plant Mol. Biol. 15:207-223(1990).  
DR EMBL: X52492; CA36733.1; -  
DR InterPro: IPR002965; P\_rich\_extensn.  
DR InterPro: IPR000684; RNA\_polII\_repeat.  
DR InterPro: IPR002879; RNA\_pol\_A2.  
DR Pfam: PF01854; RNA\_pol\_A2; 1.  
DR PRINTS: PR01217; PRICHEXTENS.  
DR PROSITE: PS00115; RNA\_POL\_II\_REPEAT; UNKNOWN\_23.  
KW DNA-directed RNA polymerase; Nucleotidyltransferase; Transferase.  
FT NON\_TER 1  
SQ SEQUENCE 625 AA; 68251 MW; DC846288DF87AF5 CRC64;  
  
Query Match 100.0%; Score 38; DB 10; Length 625;  
Best Local Similarity 100.0%; Pred. No. 22;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 YSPSPS 7  
DB 419 YSPSPS 425  
  
RESULT 5  
ID 099368 PRELIMINARY; PRT: 977 AA.  
AC 099368;  
DT 01-NOV-1996 (TReMBLrel. 01, Created)  
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)  
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)  
DE DNA-directed RNA polymerase (EC 2.7.7.6) (Fragment).  
GN RPB1-C.  
OS Glycine max (Soybean).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eucosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.  
OX NCBI\_TaxID=3847;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Dietrich M.A.;  
RL Submitted (MAR-1990) to the EMBL/GenBank/DBJ databases.

[2]  
 RN SEQUENCE OF 725-977 FROM N.A.  
 RP STRAIN-VAR.MAYNE;  
 RC MEDLINE=91355869; PubMed=2103447;  
 RA Dietrich M.A., Prenger J., Guilfoyle T.J.;  
 RT "Analysis of the genes encoding the largest subunit of RNA polymerase  
 RT II in Arabidopsis and soybean";  
 RL Plant Mol. Biol. 15:207-223(1990).  
 DR EMBL: X52495; CA36736.1; -;  
 DR InterPro: IPR000684; RNA\_polII\_repeat.  
 DR InterPro: IPR000722; RNA\_pol\_A.  
 DR InterPro: IPR002879; RNA\_pol\_A2.  
 DR Pfam: PF00623; RNA\_pol\_A; 1.  
 DR Pfam: PF01854; RNA\_pol\_A2; 1.  
 DR PROSITE: PS00115; RNA\_POL\_II\_REPEAT; UNKNOWN\_17.  
 KW DNA-directed RNA polymerase; Nucleotidyltransferase; Transferase.  
 FT NON\_TER 1  
 SQ SEQUENCE 977 AA; 107939 MW; B25F9AF7BA0191D CRC64;

Query Match  
 Best Local Similarity 100.0%; Score 38; DB 10; Length 977;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 YSPSPS 7  
 Db 780 YSPSPS 786

RESULT 6  
 O8SXM3 PRELIMINARY; PRT; 1429 AA.  
 ID O8SXM3  
 AC O8SXM3  
 DT 01-JUN-2002 (TREMBlrel. 21, Created)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE LD43558P.  
 DE RPI215.  
 GN Drosophila melanogaster (Fruit fly).  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BERKELEY.  
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,  
 RA Champagne M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,  
 RA George R., Gonzalez M., Guarin H., Krommiller B., Li P., Liao G.,  
 RA Miranda A., Mungall C.J., Nunco J., Pacleb J., Paragas V., Park S.,  
 RA Patel S., Phouanavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,  
 RA Celniker S.;  
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AY089555; AAL90293.1; -;  
 SQ SEQUENCE 1429 AA; 157524 MW; 540A045520CAED66 CRC64;

Query Match  
 Best Local Similarity 100.0%; Score 38; DB 5; Length 1429;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 YSPSPS 7  
 Db 1227 YSPSPS 1233

RESULT 7  
 O00820 PRELIMINARY; PRT; 1569 AA.  
 ID O00820  
 AC O00820;  
 DT 01-JUL-1997 (TREMBlrel. 04, Created)  
 DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE RNA polymerase II largest subunit (Fragment).  
 GN RPB1.

OS Acanthamoeba castellanii (Amoeba).  
 OC Eukaryota; Acanthamoebidae; Acanthamoeba.  
 OX NCBI\_TaxID=5735;  
 RN [1]  
 RP SEQUENCE OF 1-1014 FROM N.A.  
 RX MEDLINE=97272256; PubMed=9114022;  
 RA Stillier J.W., Hall B.D.;  
 RT "The origin of red algae: implications for plastid evolution.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 94:4520-4525(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Stillier J.W., Hall B.D.;  
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: U90211; AAC18417.1; -;  
 DR InterPro: IPR002965; P\_rich\_extensn.  
 DR InterPro: IPR000684; RNA\_polII\_extensn.  
 DR InterPro: IPR000722; RNA\_pol\_A.  
 DR InterPro: IPR002879; RNA\_pol\_A2.  
 DR Pfam: PF00623; RNA\_pol\_A; 1.  
 DR Pfam: PF01854; RNA\_pol\_A2; 1.  
 DR PRINTS: PRO1217; PRICHEXTENS.  
 DR PROSITE: PS00115; RNA\_POL\_II\_REPEAT; UNKNOWN\_7.  
 FT NON\_TER 1  
 FT NON\_TER 1  
 SQ SEQUENCE 1569 AA; 174494 MW; 13CEC59063ABC9B9 CRC64;

Query Match  
 Best Local Similarity 100.0%; Score 38; DB 5; Length 1569;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 YSPSPS 7  
 Db 1538 YSPSPS 1544

RESULT 8  
 O8SSC4 PRELIMINARY; PRT; 1599 AA.  
 ID O8SSC4  
 AC O8SSC4  
 DT 01-JUN-2002 (TREMBlrel. 21, Created)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE DNA-directed RNA polymerase II.  
 DE ECU03\_0290.  
 GN Encephalitozoon cuniculi.  
 OS Encephalitozoon cuniculi.  
 OC Eukaryota; Microsporidia; Unikaryoniidae; Encephalitozoon.  
 OX NCBI\_TaxID=6035;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GB-M1;  
 RA Genoscope;  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=GB-M1;  
 RX MEDLINE=21576510; PubMed=11719806;  
 RA Katinka M.D., Duprat S., Cornillot E., Metenier G., Thomarat F.,  
 RA Prensier G., Barbe V., Peyretailade E., Brothier P., Wincker P.,  
 RA Delbac F., El Alaoui H., Peyret P., Saurin W., Gouy M.,  
 RA Weissenbach J., Vivares C.P.;  
 RT "Genome sequence and gene compaction of the eukaryote parasite  
 RT Encephalitozoon cuniculi.";  
 RL Nature 414:450-453(2001).  
 DR EMBL: AL590443; CAD26175.1; -;  
 SQ SEQUENCE 1599 AA; 177715 MW; 208FA6AE3D6DCD8 CRC64;

Query Match  
 Best Local Similarity 100.0%; Score 38; DB 5; Length 1599;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 YSPSPS 7  
 Db 1473 YSPSPS 1479

## RESULT 9

09C428 PRELIMINARY; PRT; 1603 AA.  
 AC 09C428;  
 DT 01-JUN-2001 (TREMBlrel. 17, Created)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE DNA-dependent RNA polymerase II largest subunit RPB1 (Fragment).  
 GN RPB1.  
 OS Monoblepharis macrandra.  
 OC Eukaryota; Fungi; Chytridiomycota; Monoblepharidales;  
 OC Monoblepharidaceae; Monoblepharis.  
 OX NCBI\_TaxID=150278;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21336853; PubMed=11443356;  
 RA Stillier J.W., Riley J., Hall B.D.;  
 RT "Are red algae plants? A critical evaluation of three key molecular data sets."  
 RL J. Mol. Evol. 52:527-539(2001).  
 DR EMBL: AF315822; AK00312.1; -.  
 DR InterPro: IPR000901; CPase.  
 DR InterPro: IPR002965; P\_rich\_extensn.  
 DR InterPro: IPR000684; RNA\_polII\_repeat.  
 DR InterPro: IPR000722; RNA\_pol\_A.  
 DR InterPro: IPR002879; RNA\_pol\_A2.  
 DR Pfam: PF00623; RNA\_pol\_A; 1.  
 DR Pfam: PF01854; RNA\_pol\_A2; 1.  
 DR PRINTS: PRO1217; PRICHEXTENSN.  
 DR PROSITE: PS00867; CPASE\_2; UNKNOWN\_1.  
 DR PROSITE: PS00115; RNA\_POL\_II\_REPEAT; UNKNOWN\_9.  
 FT NON\_TER 1 1  
 FT NON\_TER 1603 1603  
 SO SEQUENCE 1603 AA; 177880 MW; E0184E2DEB2A602 CRC64;  
 Query Match 100.0%; Score 38; DB 3; Length 1603;  
 Best Local Similarity 100.0%; Pred. No. 59;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 YSPSPS 7  
 DB 1538 YSPSPS 1544  
 RESULT 10  
 096446 PRELIMINARY; PRT; 1605 AA.  
 ID 096446;  
 AC 096446;  
 DT 01-MAY-1999 (TREMBlrel. 10, Created)  
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE Largest subunit of RNA polymerase II.  
 GN RPB1.  
 OS Vallimorpha necatrix.  
 OC Eukaryota; Microsporidia; Burenellidae; Vallimorpha.  
 OX NCBI\_TaxID=6039;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99110933; PubMed=9892676;  
 RA Hitt R.P., Logsdon J.M. Jr., Healy B., Dorey M.W., Doilittle W.F.,  
 RA Embley T.M.;  
 RT "Microsporidia are related to Fungi: evidence from the largest subunit of RNA polymerase II and other proteins."  
 RL Proc. Natl. Acad. Sci. U.S.A. 96:580-585(1999).  
 DR EMBL: AF060234; AAD12604.1; -.  
 DR InterPro: IPR002965; P\_rich\_extensn.  
 DR InterPro: IPR000684; RNA\_polII\_repeat.  
 DR InterPro: IPR000722; RNA\_pol\_A.  
 DR InterPro: IPR002879; RNA\_pol\_A2.  
 DR Pfam: PF00623; RNA\_pol\_A; 1.  
 DR Pfam: PF01854; RNA\_pol\_A2; 1.

DR PRINTS: PRO1217; PRICHEXTENSN.  
 DR PROSITE: PS00115; RNA\_POL\_II\_REPEAT; UNKNOWN\_17.  
 SQ SEQUENCE 1605 AA; 180946 MW; EDC20410F5288A2C CRC64;  
 Query Match 100.0%; Score 38; DB 5; Length 1605;  
 Best Local Similarity 100.0%; Pred. No. 59;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 YSPSPS 7  
 DB 1509 YSPSPS 1515

## RESULT 11

004142 PRELIMINARY; PRT; 1685 AA.  
 ID 004142;  
 AC 004142;  
 DT 01-JUL-1997 (TREMBlrel. 04, Created)  
 DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE RNA polymerase II largest subunit (Fragment).  
 GN RPB1.  
 OS Porphyra yezoensis.  
 OC Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae; Porphyra.  
 OX NCBI\_TaxID=2788;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-US1;  
 RX MEDLINE=97272256; PubMed=9114022;  
 RA Stillier J.W., Hall B.D.;  
 RT "The origin of red algae: implications for plastid evolution."  
 RL Proc. Natl. Acad. Sci. U.S.A. 94:4520-4525(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-US1;  
 RA Stillier J.W., Hall B.D.;  
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: U90208; AAC17924.1; -.  
 DR HSSP: Q9KWU6; IHOM.  
 DR InterPro: IPR000292; For/Mit.transpt.  
 DR InterPro: IPR000684; RNA\_polII\_repeat.  
 DR InterPro: IPR000722; RNA\_pol\_A.  
 DR InterPro: IPR002879; RNA\_pol\_A2.  
 DR Pfam: PF00623; RNA\_pol\_A; 1.  
 DR Pfam: PF01854; RNA\_pol\_A2; 1.  
 DR PROSITE: PS01005; FORMATE\_NITRITE\_TP\_1; UNKNOWN\_1.  
 DR PROSITE: PS00115; RNA\_POL\_II\_REPEAT; UNKNOWN\_3.  
 FT NON\_TER 1 1  
 FT NON\_TER 1685 1685  
 SO SEQUENCE 1685 AA; 183735 MW; 15069F866EAE021D CRC64;  
 Query Match 100.0%; Score 38; DB 10; Length 1685;  
 Best Local Similarity 100.0%; Pred. No. 63;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 YSPSPS 7  
 DB 1612 YSPSPS 1618  
 RESULT 12  
 09FRB0 PRELIMINARY; PRT; 1741 AA.  
 ID 09FRB0;  
 AC 09FRB0;  
 DT 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Similar to Arabidopsis thaliana DNA-directed RNA polymerase (EC 2.7.7.6) II largest chain (JDM01).  
 GN Oryza sativa (rice).  
 OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzaceae; Oryza.



OX NCBI\_TaxID=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Chow T.-Y., Hsing Y.-I.C., Chen H.-H., Wu H.-P., Chao Y.-T.,  
 RA Liu S.-M., Hsiao Y.-Y., Huang J.-J., Lee P.-F., Su C.-L., Chen C.-S.,  
 RA Shaw J.-F.;  
 RT "Oryza sativa PAC P0001A07 genomics sequence";  
 RL Submitted (MAR-2001) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AC084218; AAC48836.1; -;  
 DR InterPro; IPR002965; P\_rich\_extensn.  
 DR InterPro; IPR000684; RNA\_polII\_repeat.  
 DR InterPro; IPR000722; RNA\_pol\_A.  
 DR Pfam; PF00623; RNA\_pol\_A; 1.  
 DR Pfam; PF01854; RNA\_pol\_A2; 1.  
 DR PRINTS; PR01217; PRICHEXTENS.  
 DR PROSITE; PS00115; RNA\_POL\_II\_REPEAT; UNKNOWN\_13.  
 KW DNA-directed RNA polymerase.  
 SQ SEQUENCE 1741 AA; 194123 MW; 207D87745CD94F1C CRC64;

Query Match 100.0%; Score 38; DB 10; Length 1741;  
 Best Local Similarity 100.0%; Pred. No. 65;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 YSPTSPS 7  
 DB 1600 YSPTSPS 1606

RESULT 13  
 O9BMP3 PRELIMINARY; PRT; 1743 AA.

AC O9BMP3;  
 DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE DNA-dependent RNA polymerase II largest subunit RPB1 (Fragment).  
 GN RPB1.  
 OS Monosiga brevicollis.  
 OC Eukaryota; Chaetognathida; Codonosigidae; Monosiga.  
 OX NCBI\_TaxID=81824;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21336853; PubMed=11443356;  
 RA Stiller J.W., Riley J., Hall B.D.;  
 RT "Are red algae plants? A critical evaluation of three key molecular  
 RT data sets.";  
 RL J. Mol. Evol. 52:527-539(2001).  
 DR EMBL; AF315821; AAK0311.1; -;  
 DR InterPro; IPR002965; P\_rich\_extensn.  
 DR InterPro; IPR000684; RNA\_polII\_repeat.  
 DR InterPro; IPR000722; RNA\_pol\_A.  
 DR InterPro; IPR002879; RNA\_pol\_A2.  
 DR Pfam; PF00623; RNA\_pol\_A; 1.  
 DR Pfam; PF01854; RNA\_pol\_A2; 1.  
 DR PRINTS; PR01217; PRICHEXTENS.  
 DR PROSITE; PS00115; RNA\_POL\_II\_REPEAT; UNKNOWN\_20.  
 RT NON\_TER  
 FT SEQUENCE 1743 AA; 191182 MW; 23D684B24A6D63B2 CRC64;

Query Match 100.0%; Score 38; DB 5; Length 1743;  
 Best Local Similarity 100.0%; Pred. No. 65;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 YSPTSPS 7  
 DB 1512 YSPTSPS 1518

RESULT 14  
 O9AXN0 PRELIMINARY; PRT; 1746 AA.  
 AC O9AXN0;

DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE DNA-dependent RNA polymerase II largest subunit RPB1.  
 GN RPB1.  
 OS Glaucosphaera vacuolata.  
 OC Eukaryota; Glaucocystophyceae; Gloeochaetales; Glaucosphaeraceae;  
 OC Glaucosphaera.  
 OX NCBI\_TaxID=38265;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21336853; PubMed=11443356;  
 RA Stiller J.W., Riley J., Hall B.D.;  
 RT "Are red algae plants? A critical evaluation of three key molecular  
 RT data sets.";  
 RL J. Mol. Evol. 52:527-539(2001).  
 DR EMBL; AF315820; AAK0310.1; -;  
 DR InterPro; IPR002965; P\_rich\_extensn.  
 DR InterPro; IPR000684; RNA\_polII\_repeat.  
 DR InterPro; IPR000722; RNA\_pol\_A.  
 DR InterPro; IPR002879; RNA\_pol\_A2.  
 DR Pfam; PF00623; RNA\_pol\_A; 1.  
 DR Pfam; PF01854; RNA\_pol\_A2; 1.  
 DR PRINTS; PR01217; PRICHEXTENS.  
 DR PROSITE; PS00115; RNA\_POL\_II\_REPEAT; UNKNOWN\_12.  
 SQ SEQUENCE 1746 AA; 193767 MW; C0769A3392D6746D CRC64;

Query Match 100.0%; Score 38; DB 10; Length 1746;  
 Best Local Similarity 100.0%; Pred. No. 65;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 YSPTSPS 7  
 DB 1606 YSPTSPS 1612

RESULT 15  
 O9HGZ5 PRELIMINARY; PRT; 1748 AA.

AC O9HGZ5;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE RNA polymerase II largest subunit.  
 GN RPB1.  
 OS Aspergillus oryzae.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 OC Eurotiiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
 OX NCBI\_TaxID=5062;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RI840;  
 RA Nakajima K., Chang Y., Suzuki T., Jigami Y., Machida M.;  
 RT "Molecular cloning and characterization of rpbA encoding RNA  
 RT polymerase II largest subunit from filamentous fungus, Aspergillus  
 RT oryzae.";  
 RL Submitted (AUG-1998) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AB017184; BAB12227.1; -;  
 DR InterPro; IPR002965; P\_rich\_extensn.  
 DR InterPro; IPR000684; RNA\_polII\_repeat.  
 DR InterPro; IPR000722; RNA\_pol\_A.  
 DR InterPro; IPR002879; RNA\_pol\_A2.  
 DR Pfam; PF00623; RNA\_pol\_A; 1.  
 DR Pfam; PF01854; RNA\_pol\_A2; 1.  
 DR PRINTS; PR01217; PRICHEXTENS.  
 DR PROSITE; PS00115; RNA\_POL\_II\_REPEAT; UNKNOWN\_4.  
 SQ SEQUENCE 1748 AA; 193131 MW; 6F178FC12A85E5B7 CRC64;

Query Match 100.0%; Score 38; DB 3; Length 1748;  
 Best Local Similarity 100.0%; Pred. No. 65;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 YSPTSPS 7

Mon Jun 23 16:45:06 2003

us-09-733-773b-3.rspt

Page 6

Db           |||||||  
          1652 YSPTSPS 1658

Search completed: June 23, 2003, 16:18:15  
Job time : 31 secs